INDIAN AGRICULTURAL
RESEARCH INSTITUTE, NEW DELHI.

18884

I.A.R.I. 6.
MGIPC—S4—10 AR—21.6.49—1,000.

PROCEEDINGS

OF THE

ROYAL SOCIETY OF LONDON

SERIES A. MATHEMATICAL AND PHYSICAL SCIENCES

VOL CLXIX

LONDON
Printed and published for the Royal Society
By the Cambridge University Press
Bentley House, N.W. 1

7 March 1939

18884

PRINTED IN GREAT BRITAIN BY
WALTER LEWIS, M.A.
AT THE CAMBRIDGE UNIVERSITY PRESS

CONTENTS

SERIES A VOL CLXIX

No. A 936. 6 December 1938

PAGE

Address of the President, Sir William Bragg, O.M., at the Anniversary Meeting, 30 November 1938 1

The simultaneous sorption of carbon bisulphide and water vapour by activated charcoals. By J. L. Lizius and A. J. Allmand, F.R.S. . 25

The band spectrum of silicon monosulphide and its relation to the band spectra of similar molecules. By R. F. Barrow and W. Jevons. (Plates 1, 2) 45

The behaviour of polar molecules in solid paraffin wax. By R. W. Sillars 66

Ionic recombination in air. By J. Sayers 83

On the Hamilton-Jacobi theory and quantization of a dynamical continuum. By P. Weiss 102

On the Hamilton-Jacobi theory and quantization of generalized electrodynamics. By P. Weiss 119

The statistical mechanics of co-operative assemblies. By E. A. Guggenheim 134

No. A 937. 22 December 1938

The Bakerian Lecture. The structure of benzene. By C. K. Ingold, F.R.S. 149

The photography of airscrew sound waves. By W. F. Hilton. (Plates 3–8) 174

The dissociation of the ammonium ion and the basic strength of ammonia in water. By D. H. Everett and W. F. K. Wynne-Jones . . . 190

The low temperature properties of gaseous helium. By H. S. W. Massey and R. A. Buckingham 205

The kinetics of electrode reactions. I and II. By J. N. Agar and F. P. Bowden. (Plate 9) 206

The effect of solvents on the continuous absorption spectrum of bromine. By R. G. Aickin, N. S. Bayliss and A. L. G. Rees 234

PAGE

The development of the spark discharge. II. By T. E. Allibone and J. M. Meek. (Plates 10–16) 246

The excitation of γ-radiation in processes of proton capture by light elements. By S. C. Curran, P. I. Dee and V. Petržílka 269

No. A 938. 7 February 1939

Relaxation methods applied to engineering problems. IV. Problems relating to elastic stability and vibrations. By K. N. E. Bradfield, D. G. Christopherson and R. V. Southwell, F.R.S. 289

Critical and co-operative phenomena. III. A theory of melting and the structure of liquids. By J. E. Lennard-Jones, F.R.S., and A. F. Devonshire 317

Collective electron ferromagnetism. II. Energy and specific heat. By Edmund C. Stoner, F.R.S. 339

The nature of sliding and the analysis of friction. By F. P. Bowden and L. Leben. (Plates 17–24) 371

The area of contact between stationary and between moving surfaces. By F. P. Bowden and D. Tabor. (Plates 25, 26) 391

The electronic structure of some polyenes and aromatic molecules. VII. Bonds of fractional order by the molecular orbital method. By C. A. Coulson 413

The spectrum of mono-deuteromethane in photographic infra-red. I. Preparation of the gas, and measurements of a simple band at 9021 cm.$^{-1}$. By W. H. J. Childs and H. A. Jahn. (Plate 27) 428

Electron transfer in mercury vapour. By F. L. Arnot and M. B. M'Ewen 437

No. A 939. 7 March 1939

A new Coriolis perturbation in the methane spectrum. III. Intensities and optical spectrum. By W. H. J. Childs and H. A. Jahn. (Plate 28) 451

The emission band spectrum of chlorine. III. Continua. By the late W. H. B. Cameron and A. Elliott. (Plate 29) 463

PAGE

A band spectrum attributed to NBr. By A. Elliott. (Plate 30) . . 469

The general theory of relaxation methods applied to linear systems. By G. Temple 476

The absorption spectra of carbon dioxide and carbon oxysulphide in the vacuum ultra-violet. By W. C. Price and D. M. Simpson. (Plate 31) 501

Statistical theory of the adsorption of double molecules. By T. S. Chang 512

Concerning the scattering of fast electrons and of cosmic-ray particles. By E. J. Williams 531

The diffusion of radon gas mixtures. By W. Hirst and G. E. Harrison . 573

Index 587

Address of the President
Sir William Bragg, O.M., at the Anniversary Meeting, 30 November 1938

According to our honoured custom I preface this annual address by references to those of our Fellows whom death has taken from us during the past year.

ARTHUR HUTCHINSON (1866–1937), Emeritus Professor of Mineralogy in the University of Cambridge and lately Master of Pembroke College, was elected a Fellow of the Royal Society in 1922. He had served on the Council from 1932 to 1934 and was a Vice-President for the year 1933–4.

While still young Hutchinson made a name for himself by brilliant researches on the chemistry and crystallography of stokesite and other minerals. He devised a stereographic projector and a slide rule which are much used by students of crystallography: and he was skilful in the construction of crystal models and lecture-room apparatus. As a teacher in the University of Cambridge he was highly successful: though it was not until 1926 that he became Professor of Mineralogy, his work had long been of professorial standing. When the use of X-ray methods from 1914 onwards opened up a new crystal science, he at once attached himself to its ways and aided its development. He designed his instruments afresh. He devoted his energies towards the organization of his Department to meet the altered needs in teaching and research. He was a devoted servant to Science, an excellent investigator, an able teacher and one of the kindest and most loyal of friends.

GEORGE HENRY FALKINER NUTTALL was born in San Francisco in 1862. He was educated in England, France, Germany and Switzerland, and to this cosmopolitan upbringing he owed his ability to speak several languages. He returned to America in 1878 and obtained the degree of M.D. at the University of California in 1884. Nuttall's scientific work covered a very wide field; in addition to several books, he published numerous papers on bacteriology, serology, hygiene, tropical medicine and parasitology. His outstanding achievement was the discovery that the blood and other body fluids of animals possess bactericidal power, a discovery which laid the foundations of all work on serum antitoxins. In 1904 he published his classical monograph *Blood immunity and blood relationship;* this investigation has clearly demonstrated the existence of a distinct similarity in chemical structure of the blood in animals which are related phylogenetically.

Nuttall was elected the first Quick Professor of Biology in Cambridge and he held this Chair until 1931. He was elected a Fellow of the Royal Society in 1904, he was Sc.D. of Cambridge and had received honorary degrees from five other universities.

In 1901 Nuttall founded the *Journal of Hygiene* and in 1908 *Parasitology*, both of which he edited with great distinction. He died suddenly in London on 16 December 1937.

Louis Napoleon George Filon died on 29 December 1937 at the age of 62. Apart from three years at King's College, Cambridge, as one of the early "advanced students", his whole academic life was spent at University College, London. His contributions to applied mathematics were in the classical field and particularly in the theory of elasticity. His patience and thoroughness and his skill in algebraic manipulation qualified him well for work in this field, and his physical insight led him to choose problems of importance and significance. His most notable contributions were the principle of "generalized plane stress" which he enunciated early in his career, and the subsequent application of this principle in the development of photo-elasticity as a practical means of expressing stress distribution. His researches are contained in nearly sixty books and papers published throughout the course of a very busy life, for he was also an able and enthusiastic teacher and took a very active part in the administration of his University.

He was elected a Fellow in 1910 and served as Vice-President of the Society for 1936 and 1937.

Alfred Barton Rendle, who died in January of illness contracted on his way to attend the Jubilee Session of the Indian Science Congress, was a botanist esteemed abroad and at home for his knowledge, judgment and helpfulness to others.

When Rendle entered St John's College, Cambridge, in 1883, an interest in plants, encouraged at school, led him to study botany. He graduated in 1887 and was appointed in 1888 an Assistant in the Botanical Department of the British Museum. In a very short time the quality of his work had brought him a request from Berlin to write a monograph on a family of plants for the *Pflanzenreich*, and similar appeals on behalf of the *Index Florae Sinensis* and the *Flora of Tropical Africa* in this country. Permitted by the Trustees of the Museum to comply with these requests, Rendle showed his readiness to help others and incidentally added to the prestige of the department, to the Keepership of which he was promoted in 1906. His judgment was subjected to a more searching test, and earned correspondingly enhanced appreciation, during his service from 1905 till 1935 as an active member of the Editorial Committee of the International Laws of Botanical Nomenclature. The Royal Society, to which Rendle was elected in 1909, has reason to know that systematic botany was not the only subject to benefit from his exact knowledge and impartial judgment for, from 1913 till he died, his service on its various committees was almost continuous and he was a member of the Council during 1929–31.

Sir JAMES CRICHTON-BROWNE, who died on 31 January 1938 at the age of 97, was the oldest Fellow of the Royal Society. For many years he occupied a unique position in the profession of medicine. He followed the speciality of his father, who was a distinguished psychiatrist, and devoted the greater part of his life to the study and the welfare of the insane. At a relatively early age he was appointed to the charge of the West Riding Asylum at Wakefield and at once instituted a study of the brain and its disorders. The pathological laboratory which he established there was the first of its kind in England, and under his guidance Wakefield became one of the most important centres of neurology in this country. The Reports which he published "to give stimulus to the arrangement and elaboration of material hitherto buried in case books" contained important papers by Ferrier, Hughlings Jackson, Clifford Allbutt and others, who at that time had no more suitable medium of publication. Later he took a major part in the foundation of *Brain*. After he had been appointed, at the age of 35, Lord Chancellor's Visitor in Lunacy, his work was mainly administrative, but various reports, as that on the influence of early education on mental health, showed that he retained his earlier interests. He took great interest in the work of the Royal Institution of which he was Treasurer for thirty-seven years. He was elected to the Fellowship of the Society in 1883.

GEORGE ELLERY HALE (1868–1938). Hale's memory is to be cherished as that of a most noteworthy leader in the recent advances of our knowledge of the phenomena presented by the surface of the sun. By his invention of the spectroheliograph in the early years of his youthful studies, he opened up a new method of recording photographically the contribution made by any selected monochromatic solar radiation (and hence of any chosen chemical element) to the pictures of the features of the sun's surface as scrutinized in ordinary telescopic vision.

In 1895, he undertook the initiation and direction of the Yerkes Observatory of the University of Chicago, and carried the work forward for ten years with marked success. Before he had reached the age of 27 he had gained recognition as a leader, without a trace of self-assertion. Failure of health and the desire for better physical conditions of astronomical observations led him to California, instigated by the prospect of the foundation, by the Carnegie Institute of Washington, of a mountain observatory for intensive study of the sun. On the summit of Mount Wilson he established a solar observatory with novel equipment of powerful instruments. With these he discovered definite indications of vortical motions in and about sunspots, and was led thereby to look for signs of the Zeeman effect in the broadened lines of the sunspot spectra. The success of his design of the 60 ft. tower telescope with a 30 ft. spectrograph sunk in a vertical shaft beneath the tower, led him to design the 150 ft. tower telescope with a spectrograph of 75 ft. in a shaft 80 ft. deep. With this equipment specially designed for the purpose he proceeded to detect and measure the intensity of the general magnetic field of the sun, analogous to the earth's magnetic field.

The remarkable success attained by the use of the 100 in. reflector, presented by J. D. Hooker, in studies of the remoter nebulae encouraged Hale to take a prominent part in the colossal enterprise of the 200 in. reflector, which is now under construction. It is to our deep regret that it has not been given to Hale to see its completion. Hale became one of the Foreign Members of this Society in 1909, and was awarded our Copley Medal in 1932.

EDWARD MEYRICK, who died on 30 March 1938, was a very remarkable man and exerted an influence both as a classical master at Marlborough College and as a naturalist—especially as an entomologist—far wider than he himself probably realized. During his active life of over 83 years, he became recognized as the world's authority on the microlepidoptera. He continued working until within a few days of his death and retained to the end his firmness and accuracy of hand and eye.

It was no small satisfaction to him to see some of the boys, whose scientific interests he had fostered at Marlborough, become eminent scientists of later days.

He was educated at Marlborough and at Trinity College, Cambridge, where he was a Classical Scholar. After nearly 10 years teaching work in Australia and New Zealand, where he also studied the microlepidoptera in his spare time, he returned to Marlborough as a classical master in 1887, and was the inspiring President of the Natural History Society for 25 years.

An indefatigable collector and an accurate observer from his schooldays, he found the microlepidoptera an almost unexplored field for study, and during 60 years of intensive work described some 20,000 species, besides hundreds of new genera and several new families. His magnificent collection of over 100,000 specimens, minutely labelled, together with his own excellent water-colour drawings, and a complete index in his small, clear handwriting has been presented, under his will, to the Natural History Museum.

Meyrick revolutionized the classification of the lepidoptera, basing his rearrangement on structural characters, and enunciated three principles, now known as "Meyrick's Laws", by which he was convinced the systematist must be guided.

He became a Fellow in 1904.

JAMES ERNEST MARSH, who died on 13 April 1938, aged 77 years, was associated all his life with Oxford. From 1886 onward he was an active worker in organic chemistry, notably in the field of camphor and the related terpenes. Early to grasp the principles of space isomerism, he found a virgin field in these complex ring structures, and, in the case of the camphoric acids, was the first to obtain optically active, *dextro-* and *laevo*-varieties for both the *cis* and *trans* forms of geometric isomers.

In association with his students, notably with J. A. Gardner, he made a very original and still interesting, though unsuccessful, attack on the problem of the constitution of camphor and turpentine.

Stone preservation was the subject of a highly original little memoir by Marsh in 1926. Regarding the decay as due most frequently to bacterial life in the stone, his remedy was based on sterilization by weak alkalies. An occasional good wash with soap and water he regarded as beneficial for the outsides of buildings as of men.

In his *Origins and growth of chemical science* (1929) he stressed the claims of the British workers, Black and Boyle, on salt formation and the fixation of gases, as the real founders of a science of chemistry.

Marsh was a keen yachtsman and good company both in work and leisure. His original and independent outlook will be remembered affectionately by the many who came under his influence at Oxford.

The Cavendish family has always been conspicuous for the importance of its contribution and great devotion to public affairs. VICTOR CHRISTIAN

William Cavendish (1868–1938), ninth Duke of Devonshire, was actuated fully by this family spirit.

Part of his early manhood, after Eton and Trinity, was spent in London offices and chambers in pursuit of that particular type of knowledge which he knew would fit him more fully for the work of administering the vast estates which it was his destiny to possess.

His political life began when, as Mr Victor Cavendish, he was returned in 1891 as member for West Derbyshire in succession to his father—a seat which he held until he succeeded to the Dukedom of Devonshire in 1908.

The Duke's interest in the north of England found one of its outlets in his Chancellorship of the University of Leeds. There he continued a family association with the original Yorkshire College of Science, founded in 1874. The first President of the college was his uncle, Lord Frederick Cavendish, in whose honour one of the original chairs of the university bears the name of Cavendish. He had been High Steward of Cambridge University from 1923.

One of the Duke's empire services was his very successful Governor-Generalship of Canada from 1916 to 1921.

After Canada post-war England claimed his unremitting attention. A year in Mr Bonar Law's government as Colonial Secretary and a close connexion with the production of the Empire Exhibition at Wembley must have taken up a very great deal of his time, and it is characteristic of his methods that he quietly took on his shoulders a very large part of the financial strain of the Wembley Exhibition as principal financial guarantor without the public of the time being aware of it.

It was in 1925 that the strain of public life took such toll of his energy that a sudden collapse compelled him to limit his public work very strictly. He was elected to the Fellowship of the Society in 1914.

Arthur Edwin Boycott, who died at the comparatively early age of 61, was an eminent pathologist and also a distinguished naturalist. He received his early scientific education at Oxford and completed his medical studies at St Thomas's Hospital. Subsequently he held pathological posts at the Lister Institute and at Guy's Hospital. He was Professor of Pathology at the University of Manchester 1912–14, and at University College Hospital, London, from 1914 until 1935, when he resigned owing to ill health.

He was a partner with Haldane in researches into the causation of miners' anaemia, of caisson disease, and of the accidents attendant on deep diving and into the pathology of mountain sickness. The practical applications of these researches have been of great advantage to industry, increased materially the depth at which divers can operate with safety and have led to a better understanding of acclimatization to high altitudes.

Boycott's best known researches in pathology are those upon the volume of the blood and regeneration of blood corpuscles in anaemias produced by various causes.

Boycott was an ardent and single-minded investigator, whose influence upon his contemporaries was good and great. He was an artist in fine expression and gave a number of stimulating addresses. As editor of the *Journal of Pathology* he rendered valuable service to British pathology.

His work in the domain of natural history was equally esteemed by zoologists. He was a recognized authority on the ecology of the British land and fresh-water molluscs, and his genetical researches into the inheritance of sinistrality in the pond snail *Limnaea peregra* published in the *Transactions* of our Society in 1928 is regarded as one of the classics of genetical research.

Boycott was elected a Fellow in 1914 and served on the Council 1929–31.

JOSEPH WILLIAM MELLOR, born at Huddersfield in 1869, spent most of his early days in New Zealand. After graduation at the University of Otago, he was awarded an 1851 Exhibition Scholarship, which took him to the University of Manchester in 1899. Under Dixon he completed a number of researches, perhaps the most important being concerned with the combination of hydrogen and chlorine. His long association with the ceramic industries began with his appointment as lecturer in pottery manufacture at Stoke-on-Trent. He was the first Principal of the North Staffordshire Technical College, built in 1914.

Mellor played a conspicuous part in the progress of clay technology in Great Britain. He raised the status of the Ceramic Society from that of a local association to that of an international institution. He was quick to realise the importance of refractory materials to the well-being of an industrial nation. In 1909 he was closely associated with the formation of the Refractories Committee of the Institution of Gas Engineers which has been responsible in Great Britain for pioneer work into the properties of refractory materials.

In 1937 he completed a monumental work, *A Comprehensive Treatise on Theoretical and Inorganic Chemistry*, in sixteen volumes. The achievement is almost beyond comprehension. He wrote every word and every reference.

Mellor was a man of great courage, ability and pertinacity. He was greatly liked: perhaps, to those who knew him, his greatest attribute was his loyalty to his work and to his friends. He became a Fellow of the Society in 1927. He died on 24 May of this year.

The official notification of the election of JOHN JACOB ABEL to the Foreign Membership of the Royal Society was delivered at his house in Baltimore on the day of his death, on 26 May, a week after he had attained the age of 81 years. He was born near Cleveland, Ohio, into a family which came originally from the German Rhineland. He graduated at the University of Michigan and subsequently continued his studies in a number of the leading medical centres of the continent of Europe. He returned to the United States to hold, in succession, the Chairs of Pharmacology at Ann Arbor and Baltimore, the latter for 39 years. Of the researches in which he was engaged throughout his long career, those on the chemistry of the hormones, from epinephrine to insulin, will probably leave the most permanent mark in the history of science. He played a leading part in establishing both biochemistry and pharmacology as separate subjects of teaching and research in the United States of America, and had a powerful influence in promoting the development there of all branches of medical knowledge by experimental research. He was busily engaged in new researches, undertaken in his retirement, right up to the onset of his last brief illness.

WILLIAM ARTHUR BONE (1871–1938), a man of great energy and forceful character, made many important contributions to the chemistry of combustion, the constitution of coal and the analysis of reactions in the blast furnace. His work was characterized by great thoroughness, pertinacity, and experimental care, and was rewarded by the discovery of many interesting new features in the behaviour of gaseous mixtures on combustion. His early work on the slow combustion of hydrocarbons (1902–6) led him to hold tenaciously the theory that their oxidation occurred through successive stages of hydroxylation, to which view his experiments on the high temperature combustion of the simpler hydrocarbons lent further support. As Livesey Professor at Leeds University, he developed the technical applications of surface combustion in collaboration with the late C. D. McCourt and was awarded the Howard Potts Gold Medal of the Franklin Institute in recognition of the advance in technology so made.

As Professor of Chemical Technology at the Imperial College of Science, he built up a school of research, and with the assistance of his collaborators studied combustion at high pressures, reaching an explosion pressure as high as 10,000 atm. He discovered phenomena relating to the effect of nitrogen in carbon monoxide explosions and proved that carbon monoxide could burn even when completely dry. He studied the initiation of detonation, the helical propagation of the head of detonation, and the effect of electrical fields on the propagation of flame. He started a series of researches on the constitution of coal in which field he became the foremost investigator. He proved the benzenoid character of the coal substance. He returned in later years to his studies on the slow combustion of hydrocarbons, publishing a series of valuable papers. He was elected to the Society in 1905, gave the Bakerian Lecture in 1932, and in 1936 received the Davy Medal.

William Wallace Campbell's outstanding contribution to astronomy is his determination of the radial velocities of the stars. In 1896, when the radial velocities of only a few of the brightest stars were known, he designed a suitable spectrograph for the great telescope of the Lick Observatory, with which results of surprising accuracy were obtained. He embarked on the ambitious programme of determining the radial velocities of 2770 stars of visual magnitude greater than 5^{m} 50, and carried this through. It involved an additional observatory in Chile under Dr Wright's direction with a similar spectrograph on a 36-in. reflector.

The observations disclosed a large number of spectroscopic binaries. The peculiarities of the orbits of these stars threw light on the evolution of double stars. In addition they showed that the stars of earliest spectral type had large masses.

The radial velocities were naturally discussed to give the motion of the sun among the stars. The direction agreed substantially with that found from proper motions, but the velocity of 19 km. per sec. was determined with much greater precision. Incidentally Campbell made the discovery that the early type or hottest stars had an apparent recessional velocity of 4 km. per sec.

Campbell made six expeditions to observe solar eclipses and devised a beautiful method of showing the reversal of the Fraunhofer lines. In 1922 with Trumpler he confirmed the verification of Einstein's prediction made by the British observers in 1919.

He was elected a Fellow in 1918; he died in 1938 at the age of 56.

Alfred Edwin Howard Tutton, who died on 14 July 1938, was elected a Fellow of the Society in 1899 at the age of 34. His scientific life was mainly devoted to a single research begun when he was a young lecturer in chemistry under Sir Edward Thorpe, and continued for 40 years. For 25 years his scientific work was carried on concurrently with the duties of H.M. Inspector of Technical Schools in London, Oxford and Devonshire successively. In each of these centres he established a private laboratory in which he determined the changes in form and physical properties which are produced in the crystals of an allied series of compounds by the replacement of one metal by another. For this purpose he investigated a vast number of crystals of sulphates and selenates and their double salts. It was a striking example of persistence, patience and successful achievement.

After his retirement from the inspectorship in 1924 he returned to another task begun many years previously and involving the greatest patience and accuracy, namely, the evaluation of the length of the imperial standard in terms of wave-lengths of light.

Crystallography was the passion of his life. He published a valuable and comprehensive treatise on the subject. He also published some lighter and more popular work on ice and snow, to the study of which he devoted much time during holidays spent in the Swiss Alps.

All Tutton's work was characterized by great experimental skill and fine finish.

In the death of ERNEST WILLIAM BROWN (1866–1938) we have lost one who embodied the spirit of the older mathematical astronomy, that concerned with the motions of the planets. His name will stand as the culmination of a list, beginning with Newton and Laplace, of those who have attempted to explain the detailed motion of the moon. It is a problem that he mastered completely, though his work involved an almost unbelievable degree of elaboration including the evaluation of some ten thousand separate terms and the rejection of a still greater number. His Lunar Tables are now universally accepted as complete for the time being, and the fact that there are still minute discrepancies is left to provide future workers with interesting fields for speculation. In his later years he attacked with much success problems connected with the very peculiar motions of some of the minor planets. Brown was by birth a Yorkshireman. He learnt his subject at Cambridge, but soon afterwards removed to America. It was at Haverford that he began his great work on the moon and it was concluded when he later removed to Yale, where the remainder of his life was spent. He had the very rare distinction of being a full member both of our Society and of the National Academy of Sciences.

BASIL MOTT, born on 16 September 1859, died on 7 September of this year. When a young man Mott was an assistant to Greathead during the construction of the City and Southwark Subway, which formed the first part of the City and South London Railway. This work involved the pioneering use of the Greathead shield. So much were his skill and energy relied on that he became Greathead's partner; and he operated the railway for the first eighteen months after it was opened in December 1890. He afterwards joined Sir Benjamin Baker in the construction of the Central London Railway, which was opened in 1900. These works and others which followed were the means of establishing the practice of deep tube tunnelling as an engineering method, and the process was thereafter applied in many other cities throughout the world.

Mott was responsible, at first in partnership with Sir Benjamin Baker, and later in sole charge, for the widening of the Blackfriars Bridge completed in 1909. In later years with his partners in the firm of Mott, Hay and Anderson he was responsible for a number of important bridges. Possibly the greatest of the works with which he was associated, and one which fully absorbed his time and his energies in his later life, was the Mersey Tunnel, the largest under-water tunnel in the world, completed in 1934. In 1925 he was chairman of a committee of engineers and architects which took the necessary steps for the preservation and safety of St Paul's Cathedral.

He was elected to the Royal Society in 1927.

The Society records with regret the death of H.R.H. PRINCE ARTHUR OF CONNAUGHT, K.G., which took place on 12 September 1938. Prince Arthur was elected a Fellow on 26 March 1914 under the Statute providing for the election of Princes of the Blood Royal: he attended a meeting of the Society and signed the Charter Book on 25 June of the same year.

ROBERT LUDWIG MOND (1867–1938) was the second son of Dr Ludwig Mond, F.R.S. Like his father he was devoted to scientific research and particularly to the advancement of industrial chemistry. When young he assisted in the investigations leading to improvements in the production of zinc by the electrolysis of zinc chloride and to the discovery of nickel carbonyl. This and later work on other carbonyls were notable achievements. Yet his greatest contributions to the advancement of science were due to the enthusiasm, good judgment and generosity with which he promoted other work than his own. Under his father's early direction he designed and equipped the Davy Faraday Laboratory of the Royal Institution, and afterwards gave liberally to its support. He took a great interest in the foundation of the Maison de la Chimie in Paris, contributed largely to its funds, and served as a member of its Council He made generous gifts to the Norman Lockyer Observatory at Sidmouth. He carried on research into various problems of physics, chemistry and agriculture in his laboratory at Combe Bank.

Robert Mond did much for the archaeology of Egypt and Palestine. His discoveries during 30 years' digging in Egypt now enrich many museums; the treasures to be found in his own collection at his house in Cavendish Square make it one of the finest in private ownership.

The value of his scientific and archaeological work was acknowledged by the distinctions which he received from learned bodies in many different parts of the world. He was elected to our Fellowship a few months ago.

The past year has included many events which might well be mentioned in this annual address. One of them has, however, elbowed out most of the rest. The two numbers of *Notes and Records* which have begun a new enterprise contain interesting and informative accounts of our doings since last November. I believe that Fellows have appreciated this social and intimate addition to the publications of the Society, and that they will join with me in an expression of our gratitude to the officers who have produced it, and especially to the one who has on this, as on so many occasions, taken the lion's share of the work. The doings of the Society are noted and recorded in this periodical more effectively than I could do in a Presidential address.

I must now speak of an event which is surely uppermost in the minds of all who have received the nominations of Council for next year. We are to lose the services of Sir Frank Smith. You may be surprised at his decision not to seek renomination when we would gladly have kept him for one more year. It happens, however, that two more of our officers must step down at the end of next year, and Sir Frank has pointed out that if three were to leave at one time the business of the Society might be seriously affected. We cannot, therefore, ask him to stay. We must take advantage of the opportunity thus given us to tell him how grateful we are for the work he has done for the Society. His secretaryship has been distinguished by a rare exhibition of capable service. The place which the Society fills in public life, and its harmonious co-operation with other institutions are largely due to his wisdom and tact and his power to unite workers in a common effort. Indeed, it is impossible to think of what he has done for us, without reflecting at the same time that he has exercised his talents in a wider field, and that in the words that may seem simple but are full of meaning, he has been and is still I am glad to say a great public servant.

I should like to add a reference to one very interesting matter. As the result of a discussion between Dr Krüss, head of the Prussian State Library in Berlin, and myself, when we attended the International Documentation

Conference at Oxford in September last—a discussion initiated by Dr Krüss—I have lately received a letter from Dr Bosch, President of the Kaiser Wilhelm Gesellschaft, inviting the co-operation of the Royal Society in some scientific enterprise which would advance science and, at the same time, promote understanding and good will. Dr Bosch asks that a few representatives from this side should be the guests of the Kaiser Wilhelm Gesellschaft for a week during this winter and hopes that we would play the part of hosts in return. During these interchanged visits, lectures would be given and consultations would be held; and it is reasonable to hope that some plan of a joint work would emerge. I feel sure that this gesture of friendliness will receive a warm welcome from the Fellows of the Society.

I will now say a few words on a subject which seems to me to exhibit a greater importance the more closely it is examined.

In recent years the very great increase in the output of the results of research has placed our scientific societies in a difficult position. There is correspondingly more to be published, and, at the same time, publication costs have increased. Consequently, the societies' finances are overstrained, and publication of worthy material is restricted. The Royal Society receives a Government Grant for publication purposes, far the greatest part of which is passed on to other Societies. The grant is sufficient to meet only a portion of the applications for assistance that are made.

There is another reason why the increase in output presses hardly on our societies. In order that a proper standard of merit shall be maintained, it is customary to submit each paper, when it is sent in, to one or more referees. Now the number of possible referees is limited, and is even becoming relatively smaller as specialization increases. I may take this opportunity of expressing the gratitude of the Society to those Fellows who spend so much care on the examination of the papers submitted to them. The Officers realize regretfully that they ask them for a considerable fraction of their time and energy.

Our shelves display the effects of increase. As we look at the long rows of bound periodicals in our studies, we realize not only that many years have slipped past since we began to collect them, but also that their very appearance is an item of history. A long time ago they were comparatively small and thin. But they grew: they shrank during the bad years of the war, but soon resumed their increase, and of late they have swollen till they have become unmanageable and must needs be divided. Still the papers come pouring in, and the rate of flow even increases. It is all to the good, and a healthy growth which we do not want to check; but there are practical difficulties which invite an attempt to solve them.

It seems to me that when we try to consider this matter from points of view all round it, we are driven to seek for a precise answer to the questions, "Why *do* we publish? Why do we submit papers and why does the Society print them, if they are good enough?"

When a man submits a paper to the Society he is, in the first place and quite rightly, anxious for the satisfaction of showing what he has done to those who will understand it. Another reason which has certainly grown in strength of recent years is that he wants to establish his reputation and position. Doubtless, he has also the wish that his work may be of service, though this desire may be relatively less obvious even to himself. To the Society the opportunity for service is the principal reason; for it cannot be held that the Society exists in these days, only or even mainly, for the pleasure and profit of its Fellows.

The existence of our Society depends on our belief that knowledge is to be obtained by experiment. In that belief our founders of nearly three hundred years ago organized the rules and procedure on which we have acted ever since. Their foresight has been amply justified; but they could not have imagined how far experiment would carry their successors and others who have held the same belief. We have been led to the discovery of a natural world of vast extent; and without limits that we can see. Every part of it is of interest and every item of knowledge that we find there has a bearing on what we think and do. As we see and know, the results of the exploration have already been immense; modern life has been profoundly affected thereby. The natural knowledge which the Royal Society set out to improve has become one of the greatest influences of these times. The Society and other bodies and individuals that have followed in its footsteps have, therefore, become possessed of certain most important opportunities, which are indeed duties. One of these is still, as it has always been, the encouragement of research, wisely conceived and well directed. A second is the preparation and presentment of the results of research, so that they can be sufficiently appreciated, and can be incorporated with understanding into every activity, intellectual or physical, to which they apply. We are not, as discoverers, responsible for the uses that are made of our discoveries. We ought not to claim, in my opinion, to be given the direction of affairs in which our discoveries play a great part simply on account of that fact. But we are, at least, bound to see that our acquired knowledge is rightly stated so that it can be rightly used. We may also, as scientists, feel that we ought to help in putting knowledge to good use. This duty is actually discharged by numbers of men in these times.

It is curious to think that the feeling of responsibility is of recent growth.

There was even a time when a discovery could be considered a private possession, to be withheld from general knowledge if thought fit. Newton, Hooke, Cavendish and many others were often in no hurry to give to the world what they had found; they were influenced by indifference, by fear sometimes, even by impish playfulness, as when truths were hidden in anagrams. They did not then look upon natural knowledge as an inheritance belonging to mankind, which, when discovered, ought to be shared.

Rumford, at the end of the eighteenth century, was one of the first to accept and proclaim his responsibility, and act upon it, when he tried to apply the laws of heat to the economy of fuel. His first attempt in London was made at the Foundling Hospital. The Royal Institution, in its first form, was his attempt to give concrete expression to his idea. The founders of the British Association were urged in their work by a similar motive and were to some extent under the belief that they were repairing a defect of the Royal Society. The vigorous action of the Association in these days is a most commendable extension of its original purpose. The Royal Society undertakes a special part of the general responsibility. It is concerned with affording opportunities for the publication and discussion of original discovery, it also encourages discovery by the administration of funds entrusted to it for that purpose. And, of course, it represents officially the scientific activities of the nation, acting as an adviser on such matters in public affairs. In this field it has full opportunity for the exercise of all its powers. This limitation of service does not free the Society from the need to attend always to the one great purpose, more fully realized of recent years, that of opening to mankind the new fields of natural knowledge. The new fields are not merely to be penetrated, they are also to be made public. There is, of course, a well-known saying that praises knowledge because it is useless. This cryptic statement would seem to be an ethical injunction to the researcher; it certainly does not mean that the most desirable knowledge is of no use to any one but the discoverer.

Even within the range of activities to which the Royal Society confines itself, in which contact with the general public is not often made, regard must be paid to the future exposition of discovery. Because the results of research have become so voluminous and so important, the manner of stating them, of arranging and storing them has also become important. We have to bear this in mind when we think of the arrangement on our shelves.

At the present time, the volumes of our *Proceedings* and *Transactions* have a certain resemblance to a building site on which the contributors of materials have shot their goods: and it must be said that the shooting is often done without regard to convenience of subsequent handling. He who would construct for himself a whole building finds it difficult to make use of the materials provided. Anyone who has completed a piece of research and hopes that it may be accepted and published by the Royal Society or any Society of learning, is in the first place at pains to describe what he has done, the methods he has used and the inferences that he draws from his results. The manner of telling his tale will depend on the purpose he has in mind. He may think that he ought to set out his facts in full detail, so that any reader who is interested shall at least find all that he wants. If he is on the threshold of research work, he will often be influenced by the desire to prove his knowledge and ability; and may be led to make his tale long and complicated. It is very likely that he has not before him a mental picture of the man for whom he is writing; if he has, it is probably that of an expert, who is working in the same narrow region as himself, who knows all the technical terms and is familiar with the hidden difficulties of the subject (which, therefore, must be discussed in full), and will pick holes if he can. It is true that writing of this kind is of the highest value: expert must talk to expert.

Nevertheless, such writings make dull and difficult reading for the great majority of those who are interested in scientific discovery. How few there are who can critically study a paper of a certain type is only too well known to our secretaries and the chairmen of our Sectional Committees, who must find referees to judge its merits, and to referees themselves who accept their often laborious task. When such papers are read at a meeting of the Society, the paucity of the attendance and the lack of good discussion are obvious and distressing. In fact, it has been found well to take vast numbers of such papers as read, and even then the discussion of the few that are left is often disappointing.

It is on this account that greater stress has been laid recently on the provision of opportunities for organized discussions, in which some important subject of recent development is considered by the workers in that subject. The discussion ranges over a wider field than that of a single worker's account of his new discoveries. It is much more interesting and informing to the general body, as the very satisfactory attendances have shown. Clearly there is a desire to understand the main purpose and the principal achievements of each growing subject.

I believe that Fellows, and others, have welcomed the recent plan of publishing summaries of papers as an appendix to the *Proceedings* as soon as the papers are received, publication of the papers themselves being deferred until approval has been given in the usual way. Not merely is an earlier announcement of discovery made on this plan, but the collection of summaries presents the general trend of research to those who are not specialists.

Summaries are the expression of a natural reaction to the difficulties caused by the growing specialization of research. They represent a certain recognition of the principle that discovery must be accompanied by exposition, the principle that has been little regarded in past years, but now demands acceptance.

The ideal summary is more than a mere digest or shortened form of the paper. It differs from the paper itself in that it is addressed to a wider circle of readers which may include the experts but contains also many others who should, in fact, receive the principal attention. For this reason it may be more difficult to write than the paper itself, because it must take into account the interests and the understandings of those who will rarely read the original paper, even if they could do so, but will gladly absorb the meaning and the information of the summary, especially if there is a connexion with their own work. When the summary is written with this wider view it becomes literature; though the paper itself may be no more than a record.

Another reason for greater attention to the writing of summaries is that specialization quickly takes explorers out of sight of each other, and out of sight of the main body. It is natural for them to talk to each other only, and to forget everyone else. Consequently, it becomes difficult to incorporate their work into any general scheme. Also it is dangerous for researchers to be so far out of reach. It is natural for them to lose a sense of proportion and to imagine, each one, that his is the main line, or at least a very important line of advance. Who is to call a halt when this happens, unless the work that is going on is so far appreciated that a general opinion can be formed as to its advisability? It would, of course, be impossible to put the direction of all research into the hands of an elected committee or a dictator. Yet a certain degree of control is required, and this is best exercised by general opinion reflected in the minds of those who direct in various ways the expenditure of money on research. The only way to facilitate the growth of a general opinion which is sound, is to insist that a general account shall be given which is intelligible to a sufficient number.

These considerations tempt me to ask whether it is necessary that our *Transactions* and *Proceedings* should contain so much as they do at present. When a new departure in experiment or theory is made, or a new fact is discovered, or a new correlation, it is right that a careful and complete explanation should be given; and if this is not too long, it is excellent that it should appear in the *Proceedings*. If the novelty leads, as it often does, to a steady output of observation, extension, confirmation and illustration, with many figures and bulky tables, there must come a time when complete publication to the world becomes unnecessary and even tiresome. What the scientific world wants is a general account of progress made, so that its bearing may be clear. A mass of details in continuation is of interest to other workers engaged in the same research, but very few others want the full account. The *Proceedings* should contain papers of original discovery or fresh departure; as regards continuation papers, as they might be called, summaries properly written should be enough. Continuation papers should, of course, be preserved and at least be available on demand; still better, a few hundred full-sized photograph copies could be made. This could be done at a fraction of the cost of printing, especially if the printing is difficult, full of figures and symbols. The *Proceedings* would then be much reduced in size, and would be more handy, interesting and useful.

There is a further consideration of no small practical importance. A largely attended conference on Documentation was held a few months ago at Oxford. The subjects with which it dealt were of comparatively modern importance, but were clearly of the greatest interest in a number of places, particularly in patent offices, and in the libraries of industrial research associations. The summaries of papers in scientific journals came in for much adverse comment, comment which seemed to be entirely just. The point was that summaries were written for the expert only, and were often barely sufficient at that.

Finally, if the scientific worker is no longer to be content merely to observe and to record, but also to share his knowledge with the world in general, he needs certain qualifications and a certain education which are much less necessary in the more restricted field. He is dealing not only with facts but with men: his work enters the humanities. This is really a very great matter. It draws its importance from the tremendous consequences of the increase of the knowledge of nature which are now obvious. Lord Rayleigh has lately shown the absurdity of the charge that scientific men are responsible for the evil uses that are sometimes made of their discoveries. But this responsibility is theirs, that they shall explain what they find so that their fellow-men know how they stand. The scientist cannot be expected to see that discovery is rightly used, but no one except himself can rightly describe it. It is not to be expected that the best use will be made of what is imperfectly understood.

These are problems of tactics. They cannot be fully solved while school and college still maintain an artificial division between two forms of teaching, naming one of them science and the other "humanities". They have far more in common than is generally recognized. There must, of course, be specialization, a divergence in the later stages. But the men of different types of mind must be kept together more than at present, so that when they meet in after-life they can understand each other. It happens too often now that the administrator makes mistakes because he cannot understand his technician's advice, and the technician is incapable of expressing himself so that his administrator understands him. There ought not, in fact, to be a sharp distinction between the two. It is a fundamental point that humanities and science have joined hands in the service of mankind.

Awards of Medals, 1938

Professor NIELS BOHR is awarded the COPLEY MEDAL. He has been for many years the recognized leader of theoretical atomic physics. His early work in 1913 provided the connecting link between Planck's conception of quanta and Rutherford's of the nucleus, and made it possible for the first time to constitute a consistent theory of spectroscopy. The development of this theory owes nearly as much to Bohr as did its inception. Among other things he invented the principle of correspondence according to which quantum phenomena pass over into the classical as a limiting case, according to certain definite rules which Bohr laid down. This principle was of the greatest importance in the early days of the theory.

Besides providing a rational explanation of atomic spectra, Bohr's theory made it possible to give a satisfactory explanation of the structure of atoms, and in particular to explain the changes in properties which occur as we go from one element to the next in Mendelejeff's periodic table. In doing so it provided an explanation of the shells of electrons whose existence had been proposed by J. J. Thomson. In the development of this theory Bohr's mathematical ingenuity and his application of the theory of perturbation played a large part.

When the principles of wave mechanics were initiated by de Broglie and Heisenberg, the latter of whom was a pupil of Bohr's, Bohr took an important part in the development of the new ideas, and in transforming atomic theory to fit them. The view which is held at the present day of the relationship between waves and particles is very largely due to Bohr.

In the last few years Bohr has transformed the ideas held as to the structure of the nucleus by pointing out the intimate association of the particles which comprise it, and the necessity for treating them as part of a closely knit system more like a liquid than a gas. This fundamental idea has already had far-reaching effects, and is the basis of almost all present work.

Although his published papers are of supreme importance, it is probable that Bohr's greatest contribution to physics lies in the influence which he has exerted over an enormous proportion of the leading physicists of the world. He possesses to an extraordinary extent the ability to draw ideas from minds which would otherwise probably never have produced them, and all who have fallen under his influence are conscious of his supreme power of inspiration.

Professor ROBERT WILLIAMS WOOD is awarded the RUMFORD MEDAL. The study of physical optics owes much to Professor Wood, who has been one of the leading experimenters in this field for the past 40 years. There is hardly a branch of the subject which he has not enriched by the touch of his genius.

Before the advent of Bohr's quantum theory, when our knowledge of the structure of atoms and molecules was very meagre, he had discovered the line and continuous absorption of sodium vapour, the phenomenon of resonance radiation of gases and vapours, and the quenching of this radiation by foreign gases. These discoveries opened up rich fields of research and were of the greatest value to later workers in laying the foundations of the theory of atomic and molecular spectra.

The elucidation of the phenomenon of resonance radiation demanded the utmost experimental skill and resource. Nothing less powerful than an improvized 40 ft. focus spectrograph sufficed for his work on the remarkable resonance spectra of molecules! Even now one cannot but admire the beautiful and ingenious experiments on the independent excitation of the yellow sodium lines.

In addition to his researches on the resonance radiation of metallic and other vapours, Wood investigated their magnetic rotation and dispersion. His work on the magneto-optics of sodium vapour both in the atomic and molecular state is now classical.

More recent but belonging to the same domain of experiment are the very interesting discoveries of Wood and Ellett on the magneto-optics of resonance radiation.

Wood's mastery of technique is universally acknowledged. He has introduced many ingenious and striking devices to experimental method. These are too numerous to catalogue here, but I would mention specially his method of the production of atomic hydrogen and his observations on the spontaneous incandescence of substances in atomic hydrogen which led to the invention of the atomic hydrogen welding torch by Langmuir; his very efficient and now widely used method of observing Raman Spectra; his échelette grating which has proved to be the grating *par excellence* for the investigation of the near and far infra-red; and his pioneer use of light filters in ultra-violet and infra-red photography.

A Royal Medal is awarded to Dr Francis William Aston. Few cases are known to science in which an important development has remained so much in one man's hands as has the study of isotopes, by means of the mass spectrograph, in those of Dr Aston. Aston's attention was drawn to isotopes by his work as assistant to Sir Joseph Thomson when the latter was working on the analysis of positive rays by the parabola method. The results for neon suggested very clearly that this element contained atoms of weight 20 and 22, but other explanations were conceivable. Aston attempted to settle this question by separating the two components. Fractional distillation was carried out in an elaborate apparatus of his own devising and construction. A microbalance was devised to measure the hoped-for change of density, but the result was negative. Further attempts at separation by diffusion involving the construction of elaborate devices for repeating the diffusion process many thousand times gave inconclusive results. As is now known, this was due to inadequate mixing. After the war Aston started a new line of attack by improving the apparatus for the electromagnetic analysis of positive rays. Known as the mass spectrograph, his instrument, which depends on an ingenious focusing of the rays, was powerful enough to show that the lighter constituent of neon had a mass less than corresponded to the density of the gas as a whole, which was therefore a mixture. But the results went much further, and Aston found that not neon only but the majority of elements consisted of mixture of isotopes, in the sense already established by Soddy and others for certain products of radioactivity. This discovery profoundly altered men's views as to the nature of the ordinary chemical elements. For one thing, atomic weights as ordinarily determined were seen to be merely weighted means and not fundamental constants. By various modifications of this method Aston was able to extend his experiments to cover the great majority of the known elements, and to establish the existence of hundreds of isotopes. In addition he was able to show that, with the exception of hydrogen, all the isotopes had nearly integral atomic weights taking oxygen as 16. This result was of great importance in connexion with Rutherford's theory of the nucleus.

The later developments of Aston's work are concerned mainly with the small deviations from this whole number law. Since on the theory of relativity mass and energy are equivalent, the mass of a nucleus is a measure of its internal energy, and the small divergencies from whole numbers give the energy with which the constituent particles are bound together. Using an improved mass-spectrograph capable of an accuracy of 1 in 20,000, Aston has determined this binding energy with considerable accuracy, at least for the lighter elements. His figures have been checked in many cases by the results of atomic disintegrations, and, with similar measurements by Bainbridge, form an invaluable basis for the study of nuclear changes.

By determining photometrically the proportions of the isotopes, Aston has been able to calculate "chemical" atomic weights, which in some cases have corrected those found by the older methods. Thus both in chemistry and in physics Aston's work has been of outstanding value.

A ROYAL MEDAL is awarded to Professor RONALD AYLMER FISHER. His contributions, both to the development of the logical theory of statistical methods and to the invention of efficient tools for the use of the experimental worker, have been of outstanding importance. Before Fisher entered the field (in 1912), the work of Galton, Pearson and their immediate pupils had widely extended the connotation of statistical methods, particularly in biological research. Useful descriptive methods had been invented and reasonably adequate tests of sampling errors when samples were large (say, a hundred or more observations) were made available. As a result there was an enormous increase in the number of workers who applied quantitative methods to biological, medical and sociological problems. It may be said that in freeing the statistical methods of weaknesses obvious a quarter of a century ago, Fisher has been one of the chief contributors to the improvement of the logical basis of statistical methodology. He has been the most important contributor both to the theory and practice of small sample analysis, and to the armamentarium of statistical tools for biological and technological research. His own practical contributions to the study of genetics have been neither few nor unimportant (his treatise *The Genetical Theory of Natural Selection*, 1930, and papers in the B *Proceedings* and *Transactions* of the Society are examples), but the scope of his work has covered a much wider field of scientific research interesting both the physical and biological sides of the Society.

The DAVY MEDAL is awarded to Professor GEORGE BARGER. He has been continuously engaged in research, mainly on the constitution and synthesis of natural products, since 1902.

His two most important contributions to knowledge are linked together, namely, his work on the alkaloids of ergot and on the sympathomimetic amines.

After his isolation (with Carr) of the alkaloid ergotoxine, to which the effect of ergot on the cock's comb is due, he showed that the pressor bases from putrid meat were the products of decarboxylation of amino-acids, and included *p*-hydroxyphenylethylamine, derived in this way from tyrosine. He worked out methods for the synthesis of *p*-hydroxyphenylethylamine and allied compounds, and proved that this compound and also histamine, the product of decarboxylation of histidine, occurred in extracts of ergot.

One of the earliest systematic investigations of the relation between chemical constitution and pharmacological properties was the work which he carried out in collaboration with Dale on the sympathomimetic amines of the phenylethylamine class.

He has made considerable contributions to our knowledge of the constitution of many alkaloids, including carpaine, physostigmine and yohimbine, and also of many naturally occurring derivatives of amino-acids, such as carnosine, ergothioneine and methionine. Special mention may be made of his work, carried out jointly with Harington, on the constitution and synthesis of thyroxine.

The DARWIN MEDAL is awarded to Professor FREDERICK ORPEN BOWER. His main contributions to botany have been the intensive study of the morphology and affinities of ferns and allied plants. A series of papers published in the *Transactions* of the Royal Society and the *Annals of Botany* on the development of spore-bearing organs in the Pteridophyta is a contribution to phylogeny of great importance. In his book *Origin of a Land Flora* (1908) Bower discussed evolutionary problems with special reference to the Bryophyta and Pteridophyta : 26 years later he dealt with the same problem in *Primitive Land Plants*. His three volumes on ferns (1923–8) are a classic work on the affinities of members of this group and a most valuable source of information.

In 1930 he discussed *Size and Form in Plants*. His latest book, *Sixty years of botany in Britain*, written in his eighty-third year, is in part autobiographical. Professor Bower has consistently devoted himself to a branch of botanical science in which he has long been regarded as a leading authority. By his own researches and the judicial examination of recorded facts he has thrown light on the natural affinities of certain groups of plants.

The HUGHES MEDAL is awarded jontly to Dr JOHN DOUGLAS COCKCROFT and Dr ERNEST THOMAS SINTON WALTON. The discovery by Dr Cockcroft and Dr Walton that the transmutation of elements can be effected by means of artificially accelerated particles has opened up a new line of work of outstanding interest and importance. In previous experiments, initiated by Rutherford as long ago as 1919, the transmutation of several elements had been produced by bombarding them with the α-particles from the natural radioactive substances. In this early work it was thought that bombarding particles of very great energy were required to disrupt atomic nuclei. The development of quantum mechanical ideas to nuclear problems made it possible for the first time to see how charged particles of low energy could penetrate into an atomic nucleus, and, moreover, enabled calculations to be made of the probability of penetration. It was on the basis of such calculations that Cockcroft and Walton came to the conclusion that protons accelerated by moderate potentials, of the order of a few hundred kilovolts, should be able to effect the disintegration of the lighter elements. During 1930 attempts were made to disintegrate lithium by bombardment with protons using accelerating voltages up to 300,000 V. The results were at first negative but continued experiment, in which the apparatus was redesigned to give voltages up to 600 kV, was in 1932 rewarded with the successful disintegration of lithium by bombardment with protons. Thus for the first time an atomic transmutation was effected by means entirely under the experimenter's control.

These experiments of Cockcroft and Walton gave a tremendous impetus to investigations in nuclear physics. Their work has been taken up in many laboratories and in many countries, and it has given rise to such a mass of experimental results as to form almost a new branch of science.

The simultaneous sorption of carbon bisulphide and water vapour by activated charcoals

By J. L. Lizius and A. J. Allmand, F.R.S.

(*Received* 12 *July* 1938)

The experiments described below were carried out in the sessions 1929–30 and 1930–1, subsequently to the work on the sorption of CS_2 vapour by dry charcoals already published (Allmand and Lizius 1932), and formed part of a then projected detailed study of the effect of the prior sorption of one vapour or gas on the sorption of a second. A previous instalment of this work has already appeared (Chaplin 1934), but circumstances made it necessary to abandon the major programme.

Three charcoals were experimented on, viz. *A* (an air-activated birchwood charcoal), *B* (a chemically activated pinewood charcoal), and *C* (a steam-activated coconut charcoal), of respective packing densities 0·205, 0·298 and 0·490. They have been more fully specified elsewhere (Hand and Shiels 1928). The work comprised (*a*) the measurement at 25° C of the low-pressure isothermals of CS_2 sorbed on charcoals containing constant known amounts of sorbed H_2O, and (*b*) in the case of charcoals *A* and *B*, the measurement at 25° C of the effect of variation in the amount of sorbed H_2O on the pressure exerted by a definite charge of CS_2.

Before proceeding to these experiments, it will be well to set out certain general and special conclusions drawn from previous work.

(1) The surface of a charcoal comprises areas of widely varying adsorption potential, and is normally covered with a charge of chemically adsorbed oxygen. Whilst this charge can slowly be removed by evacuation at raised temperatures as a mixture of CO_2 and CO, the surface is cleaned up still more effectively if the charcoal is allowed to adsorb some gas or vapour. As the oxides of carbon are displaced, so the quantity of vapour adsorbed progressively increases, and definitive results for the amount of adsorption at any given pressure of vapour are only obtainable when the gas displacement has virtually ceased (Allmand and Chaplin 1930*b*). The amount of gas so displaced varies much from charcoal to charcoal, and the rate of such displacement depends considerably on the vapour which is effecting it. Charcoal *B*, a chemically activated produce, is looked on (Allmand and Chaplin 1931) as having its areas of highest potential occupied by adsorbed ZnO derived from the $ZnCl_2$ used in its activation; the oxygen held on the

areas of lower adsorption potential is removed with comparative ease during the preliminary outgassing; on the admission of vapour further gas displacement is negligible, and adsorption equilibrium rapidly sets itself up, the quantities adsorbed at low pressures being relatively small compared with other charcoals. Both charcoal *C* (steam-activated) and charcoal *A* (air-activated), when exposed to CCl_4 vapour under similar conditions, show considerable gas displacement, accompanied by a progressive increase in adsorption (Allmand and Chaplin 1930*b*). This "drift" is particularly marked with *A*; charcoal *C* not only arrives at the final equilibrium state more rapidly, but also takes up larger quantities of adsorbed CCl_4 under the same pressure conditions. A relatively high charge of adsorbed vapour accelerates the displacement of the surface oxides and the establishment of equilibrium (Allmand and Puttick 1930).

(2) If CS_2 be substituted for CCl_4, charcoal *B* behaves as before, but with *A* and *C* two differences are remarked (Allmand and Lizius 1932). The amounts of CO_2 displaced from the charcoal surface are considerably greater than is the case with CCl_4, but there is no detectable "drift"; on the contrary, a slight diminution in adsorption was observed (self-poisoning). Secondly, the quantity of sorbate taken up by charcoal *A* is now greater, and not less, than that taken up by *C*. At the time, we suggested that CS_2 was able to break up the surface oxide complex more readily than CCl_4, this being the essential process governing the degree of adsorption and therefore of drift, the actual rate of displacement of the CO_2, when once formed, from the oxidized charcoal surface being of minor importance in this connexion. Whilst still retaining the distinction between these two stages and their significance, we now are of opinion that a steric factor is involved, that CS_2 molecules can readily penetrate to regions in the charcoal structure difficultly accessible to CCl_4 molecules owing to their larger size, that charcoal *A* contains a relatively higher proportion of such regions than does *C*, and that the above differences between the reciprocal behaviour of the two charcoals and the two vapours thus find an explanation.

The results of Chaplin (1934) on the displacement of the surface oxides from charcoal by HCN, CCl_4, and the two vapours acting together, seem capable of a similar explanation, one which forms a slight modification of his own views. He found that, whereas neither CCl_4 or HCN alone could completely expel the surface oxides, this was readily brought about by the two vapours acting in conjunction. All his observations are consistent with the idea that a sufficient charge of CCl_4 destroys quickly all the accessible surface complex, that HCN, which is less powerfully adsorbed, is only partly effective in this respect, but that its molecules can penetrate easily into parts of the charcoal lattice relatively inaccessible to CCl_4 and then rapidly break up the oxide layer to the limit of their ability (CS_2 would probably have a stronger action in this last respect).

(3) Water is well known to give sorption isothermals of type very different from those of substances such as CCl_4 and CS_2 (e.g. Allmand, Hand, Manning and Shiels 1929*a*). Whereas the adsorption of the latter by charcoal over the whole pressure range can be explained satisfactorily on the basis of van der Waals' adsorption on a complex surface of varying adsorption potential, any "activation energy of adsorption" being due to the necessity of displacing the oxygen layer (Allmand and Chaplin 1931, 1932), the sorption of water is, in all probability, a complex process, comprising physical adsorption at low and capillary condensation (very important with *B*) at high pressures, together with a third, intermediate, stage, of which the exact nature is not quite clear, but which is probably the production of a continuous film of liquid on the surface of the charcoal. Down to relatively low pressures, the heat of adsorption is the same as, or very little greater than, the heat of vaporization (Allmand and others 1929*a*), but in the region of ordinary physical adsorption becomes considerably higher (Allmand and King 1930).

The occurrence of a slow "drift" is very marked when determining a water-vapour isothermal, particularly on charcoals such as *A* and *C*. It is frequently accompanied by CO_2 displacement (as in other cases, again particularly marked with charcoal *A*—Allmand and King 1930), is accelerated if pressure and/or temperature be raised (Allmand, Hand and Manning 1929*b*), and there seems every reason to attribute it to the same cause as with other vapours. We think that its unusual degree is due to the small size of the H_2O molecule, which enables the latter to penetrate to points inaccessible to molecules such as CS_2, and to the probable relatively high stability of the surface oxygen complex in such regions in respect of attack by H_2O molecules. Water which is adsorbed in this way is held with extraordinary tenacity, not being removed even on prolonged high-temperature evacuation (Allmand and others 1929*b*). Having already ascribed a similar role to HCN, we are supported in our present view by the close analogy (Allmand and Chaplin 1931) between the low-pressure isothermals of HCN and of H_2O, and by the fact that HCN can displace H_2O from a supposedly water-free charcoal under conditions where CCl_4 has no effect (Chaplin 1934). Both H_2O and HCN have small molecules and form associated liquids of high dielectric constant.

Finally, in the case of charcoal *B*, when H_2O vapour is admitted after evacuation, very marked drift is observed, but unaccompanied by CO_2

evolution. This disappears completely if the charcoal be previously extracted successively with HCl and with H_2O, thereby removing the residual $ZnCl_2$ remaining after the activation process and reducing the ash content from 3·90 to 0·12 %. The drift in this case is obviously due to the hygroscopic nature of the $ZnCl_2$ present (Allmand and King 1930; also Allmand and others 1929*b*).

We may summarize the properties of the three charcoals in respect of the two vapours concerned in the present work as follows:

Charcoal A contains a high proportion of relatively "remote" surface charged with oxide complex. Some of this is accessible to CS_2, and still more to H_2O, molecules. Both these substances can destroy and displace the surface oxide layer, the H_2O acting relatively slowly on the less easily accessible and more stable complex which the CS_2 can not reach. The "external" surface of *A* is less powerfully adsorbing than is the case with *C*.

Charcoal B contains a hygroscopic constituent, and, in addition, has its regions of high adsorption potential, whether "remote" or otherwise, already permanently occupied by a mineral constituent. The remaining surface has relatively weak adsorbing properties. Its volume contains a high proportion of pores of dimensions in which capillary condensation can occur at suitable values of p_{H_2O}.

Charcoal C resembles *A*, but has a larger accessible area of high adsorption capacity, and a smaller proportion of remote surface.

EXPERIMENTAL

The apparatus employed was essentially that used for the work on the low-pressure isothermals of CCl_4 (Chaplin 1928), incorporating the minor alterations introduced during the work on CS_2 (Allmand and Lizius 1932), and further modified by the provision of a supply bulb, shut off by a tap, which contained water freed from air by the usual method of repeated cycles of freezing—evacuation—melting. The general technique, involving the separation by fractional refrigeration of the vapour of the sorbate, of CO_2 and of CO (the two last displaced from the charcoal), and the separate measurement of their pressures by means of a Pirani gauge, has been fully described in the papers just mentioned. The simultaneous presence of H_2O and CS_2 vapours, however, with condensation temperatures nearer to one another than had previously been the case, made the problem of satisfactory separation of the gaseous phase constituents more difficult than formerly.

We finally decided to use for the separation in question the three refrigerants given in the following table:

Refrigerant	Melting-point	Vapour pressure at melting-point: Of ice mm.	Vapour pressure at melting-point: Of carbon bisulphide mm.
Chlorobenzene	−45·2°	0·0525	10·7
Chloroform	−63·5°	0·0050	2·58
Ethyl acetate	−83·6°	0·0002	0·398

Up to CS_2 pressures of about 0·3 mm., melting ethyl acetate is the most suitable agent, but at higher CS_2 pressures, one of the other refrigerants must be employed, depending on the conditions.

A description of the experimental procedure follows. References to apparatus parts are to fig. 2 of Chaplin (1928).

Charge of water constant—charge of carbon bisulphide varied. The charcoal was evacuated at 110° down to 10^{-4} mm., cooled with the pumps running, and weighed. It was then charged with H_2O (immersing the container in liquid air to save time) and the quantity taken up determined by reweighing. In a similar way it was charged with CS_2 (liquid air cooling unnecessary) and reweighed. The container, with its tap closed, was then immersed for an hour in water at a suitable temperature (25° C with large, but higher temperatures with smaller, charges of H_2O—it is undesirable for H_2O in any quantity to condense in the neck of the container). The container tap was then opened with F_1 immersed first in liquid air and then in melting methyl*cyclo*hexane, and the permanent gas and CO_2 displaced from the charcoal successively pumped off, the H_2O and CS_2 held down in F_1 being subsequently taken back on to the charcoal by immersing the container in liquid air until the Pirani gauge reading was within 0·1 V of the zero figure. For desorption points, some CS_2 and H_2O were first removed from the charcoal by means of liquid air on F_1. The H_2O was held down by melting ethyl acetate, and the CS_2 removed by the pumps or by liquid air on the second freezer F_2. Finally, the H_2O was taken back from F_1 on to the charcoal by immersing the container in liquid air. For sorption points, CS_2 was obtained from the supply, freed from air and CO_2, and then sorbed on to the charcoal by immersing the container in liquid air.

When the quantity of H_2O sorbed by the charcoal was relatively high, pressures of the order of 10 mm. were developed in the apparatus on opening the container tap for the purpose of the final pressure measurements, and it was found important for all the mercury seals (H_1, H_2, H_3) to be raised and lowered several times, in order to disengage small quantities of gases and vapours which were liable to be held between the mercury and glass surfaces, before commencing the separation of the gaseous mixture. When the CS_2 pressure was higher than the limit of the Pirani gauge (about 0·5 mm.), it was directly measured by means of a millimetre scale placed behind the clean-up charcoal mercury seal H_1, which was used as a manometer, the levels being read by a magnifying eyepiece. The accuracy of the measurements was of the order of 0·1 mm.

Charge of carbon bisulphide constant—charge of water varied. If the water were to be taken up by the charcoal first, the latter, evacuated at 110°, was successively charged as already described with (*a*) 200–300 mg./g. of H_2O, and (*b*) as much CS_2 as would give a pressure of about 0·04 mm. if the charcoal were anhydrous. After measuring the first set of equilibrium pressures, desorption of the H_2O in steps was commenced. This was done (i) by putting liquid air on F_1, thereby withdrawing both some CS_2 and some H_2O from the charcoal, (ii) by shutting off the container, replacing the liquid air on F_1 by melting ethyl acetate, opening the container to the apparatus, immersing it in liquid air and thereby taking back the CS_2 on to the charcoal. The container was then shut off, F_1 allowed to rise to room temperature, and the H_2O, after its pressure had been measured, removed by the pumps or by the clean-up charcoal in P. This whole procedure was repeated until the amount of sorbed H_2O had been reduced to the required low figure. The conditions when determining subsequent sorption points, except in that the H_2O was added step by step, were essentially the same as those in experiments when CS_2 was admitted to the charcoal before sorbing any H_2O. In this case, after the charcoal had been charged with CS_2, the displaced gases were removed, and the requisite amount of H_2O vapour admitted. The subsequent procedure was as already described, the CS_2 always being taken back on to the charcoal after any measurement of pressure.

It may be added that, from the value of p_{CS_2} found in one or two cases for the last desorption point (water nearly completely removed), it was calculated that the loss of CS_2 taking place during a series of measurements was of the order of 2 mg./g.

RESULTS

Three isotherms (1–3) were determined using charcoal A with various constant charges of sorbed H_2O, two similar isotherms (4 and 5) with charcoal B, and one (6) with charcoal C. Two experiments (7 and 8) with constant CS_2 charge and varying H_2O content were carried out with charcoal *A*, and one (9) with charcoal *B*. The essential details describing the separate experiments follow; all charges, whether of H_2O or of CS_2, are expressed as mg./g. of charcoal.

1 (fig. 1, ii). Experimental points 8. *Constant* H_2O *charge* 2·0. CS_2 *charge*: initially 79·7; desorption (3 points) to 12·9; resorption (2 points) to 76·7; second desorption (2 points) to 15·0. The CS_2 pressure found at point 3 was known to be low as, by error, the vapour space was put into communication with the clean-up charcoal at liquid-air temperature for too long (3 instead of 2 min.) when separating CO_2 and CS_2 (Allmand and Lizius 1932, p. 557). It will be seen to fall right off the curve.

2 (fig. 1, iii). Experimental points 8. *Constant* H_2O *charge* 17·2. CS_2 *charge*: initially 71·2; desorption (4 points) to 6·5; resorption (3 points) to 56·5.

3 (fig. 1, iv). Experimental points 10. *Constant* H_2O *charge* 102·6. CS_2 *charge*: initially 6·5; sorption (5 points) to 80·2; desorption (4 points) to 14·4.

4 (fig. 2, ii). Experimental points 9. *Constant* H_2O *charge* 10·2. CS_2 *charge*: initially 1·2; sorption (5 points) to 23·6; desorption (3 points) to 4·5.

5 (fig. 2, iii). Experimental points 9. *Constant* H_2O *charge* 106·9. CS_2 *charge*: initially 3·7; sorption (5 points) to 27·5; desorption (3 points) to 4·7.

6 (fig. 3, ii). Experimental points 11. *Constant* H_2O *charge* 18·6. CS_2 *charge*: initially 64·5; desorption (6 points) to 8·9; resorption (4 points) to 64·4.

7 (fig. 4, circles). Experimental points 12. *Constant* CS_2 *charge* 54·9, put on after the first charge of H_2O. H_2O *charge*: initially 256·1; desorption (11 points) to 7·1. At the four highest H_2O charges, p_{H_2O} exceeded saturation vapour pressure at room temperature, and negligible amounts of liquid water condensed in the container just below the tap. The pressure of the original CS_2 charge in absence of H_2O would have been 0·0362 mm. The pressure of 0·035 mm. obtained during the last reading indicates that a loss of about 0·8–1 mg./g. took place during the experiment.

8 (fig. 4, squares). Experimental points 14. *Constant* CS_2 *charge* 53·9, put on before the first charge of H_2O. H_2O *charge*: initially 305·9; desorption (13 points) to 4·6. As in Exp. 7, slight condensation of H_2O in the container was observed with the first five water charges. The pressure of the original CS_2 charge in absence of H_2O would have been 0·0344 mm. The pressure of 0·032 mm. observed during the last reading points to a loss of about 1·7 mg./g. during the experiment.

9 (fig. 5). Experimental points 49. *Constant* CS_2 *charge* 14·2, put on after the first H_2O charge. H_2O *charge*: initially 209·8; desorption (8 points) to

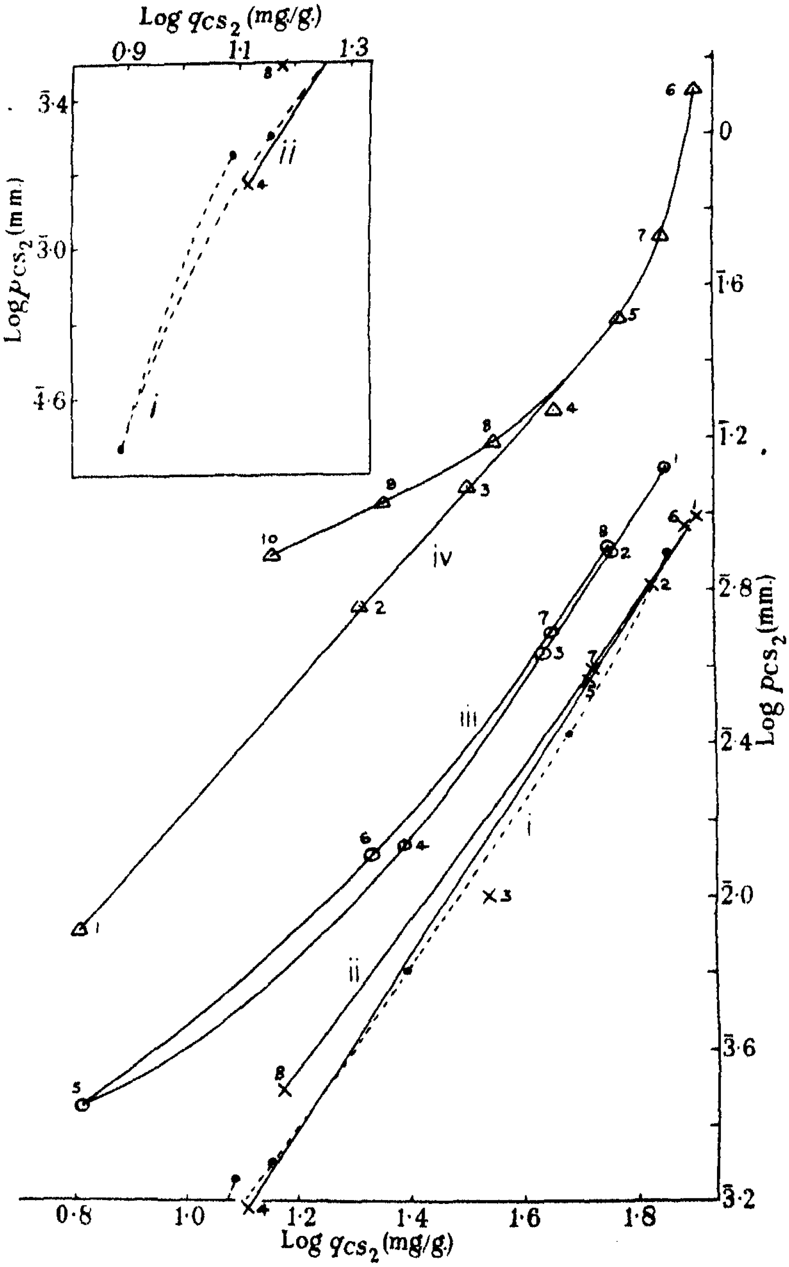

Log q_{CS_2} (mg/g)
0·9
1·1
1·3
$\bar{3}$·4
$\bar{3}$·0
$\bar{4}$·6
Log p_{CS_2} (mm.)
ii
i
0
$\bar{1}$·6
$\bar{1}$·2
$\bar{2}$·8
$\bar{2}$·4
$\bar{2}$·0
$\bar{3}$·6
$\bar{3}$·2
Log p_{CS_2} (mm.)
iv
iii
ii
i
0·8
1·0
1·2
1·4
1·6
1·8
Log q_{CS_2} (mg/g)

Fig. 1

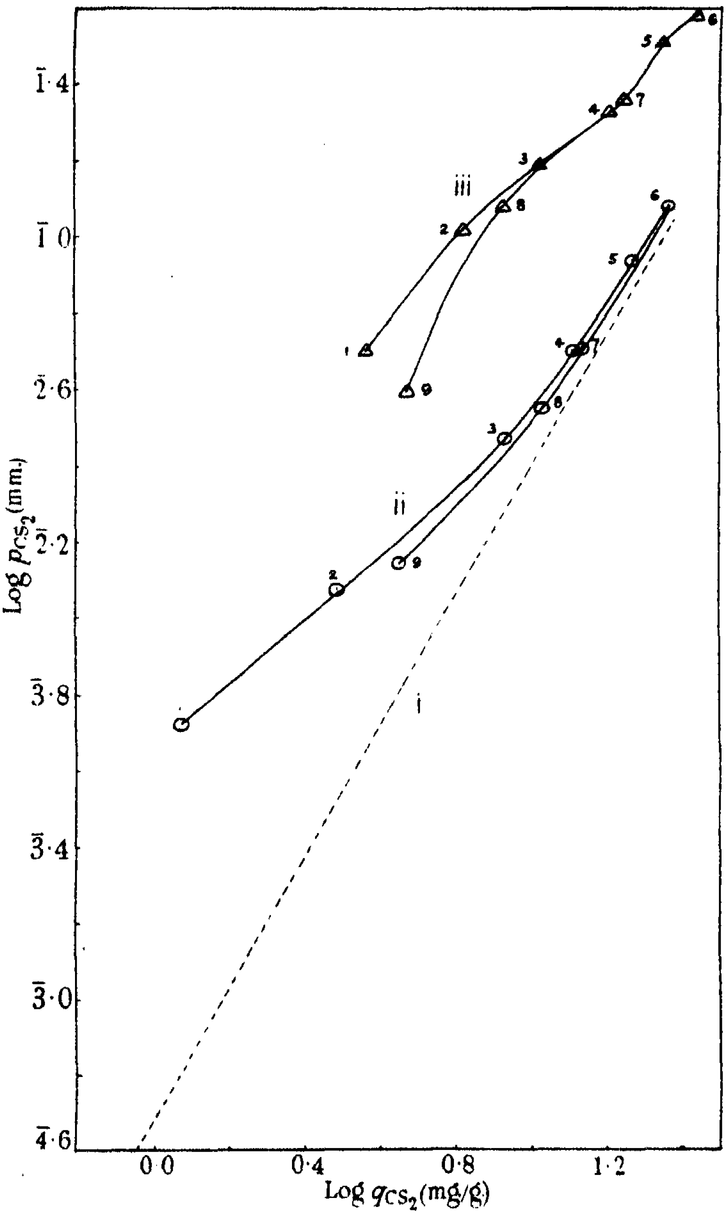

$\bar{1}$·4
$\bar{1}$·0
$\bar{2}$·6
$\bar{2}$·2
$\bar{3}$·8
$\bar{3}$·4
$\bar{3}$·0
$\bar{4}$·6
Log p_{CS_2}(mm.)
0·0
0·4
0·8
1·2
Log q_{CS_2}(mg/g)
i
ii
iii
1
2
3
4
5
6
7
8
9

FIG. 2

12·9; resorption (8 points) to 285·4; second desorption (15 points) to 21·0; second resorption (17 points) to 212·6. Slight H_2O condensation was observed during points 1–2, 13–20, and 46–49. The original CS_2 charge would have exerted a pressure, in absence of water, of 0·0447 mm. (horizontal line in fig. 5), and a pressure of 0·043 mm. was observed at point 32. At the end of the experiment, the container, with its tap open, was immersed in hot water and kept in connexion with a liquid-air trap for 2 days. At that stage, its increase in weight, over and above that at the commencement of the experiment, corresponded to 6·6 mg./g. of charcoal, and was taken as a

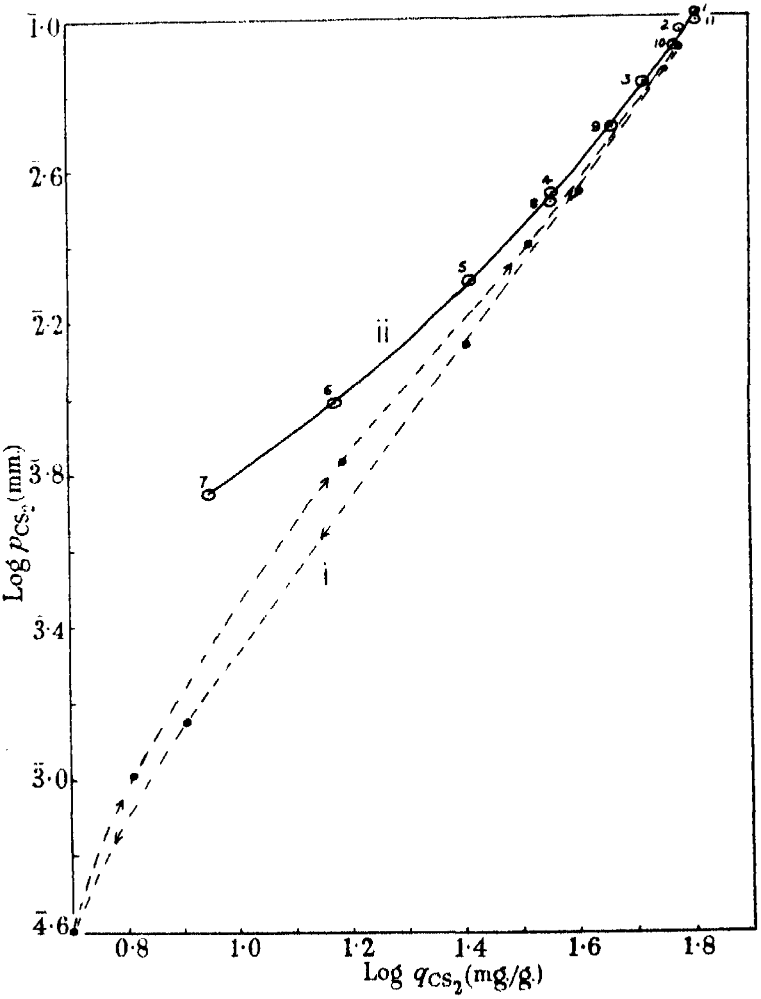


Fig. 3

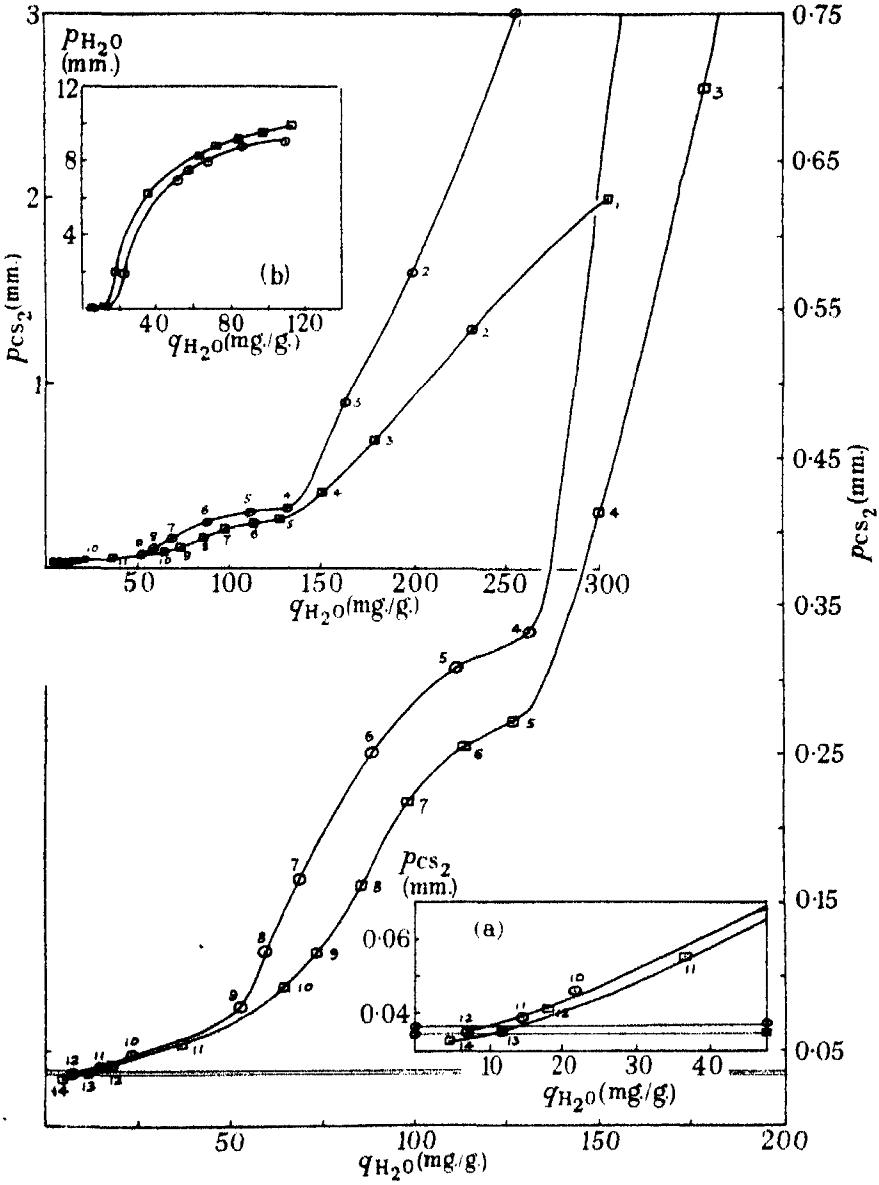

p_{CS_2}(mm.)
3
2
1
p_{CS_2}(mm.)
0·75
0·65
0·55
0·45
0·35
0·25
0·15
0·05
50
100
150
200
250
300
q_{H_2O}(mg./g.)
p_{H_2O} (mm.)
12
8
4
40
80
120
q_{H_2O}(mg./g.)
(b)
p_{CS_2} (mm.)
0·06
0·04
(a)
10
20
30
40
q_{H_2O}(mg/g)
50
100
150
200
q_{H_2O}(mg/g.)

Fig. 4

measure of the tightly bound water. After reconnecting, the liquid air on the trap was replaced by melting ethyl acetate, and the CS_2 allowed to return to the charcoal. In consequence of the considerable time required, the first separation was very incomplete, much H_2O passing back to the container. The whole process was then repeated, after discarding the residue of water remaining in the trap. The weight of the container after the second separation corresponded to a total adsorbed quantity of 18·6 mg./g. which, allowing for the 6·6 mg./g. of bound water, indicates that the original CS_2 charge of 14·6 mg./g. had fallen to 12·0 mg./g. in the course of the lengthy experiment.

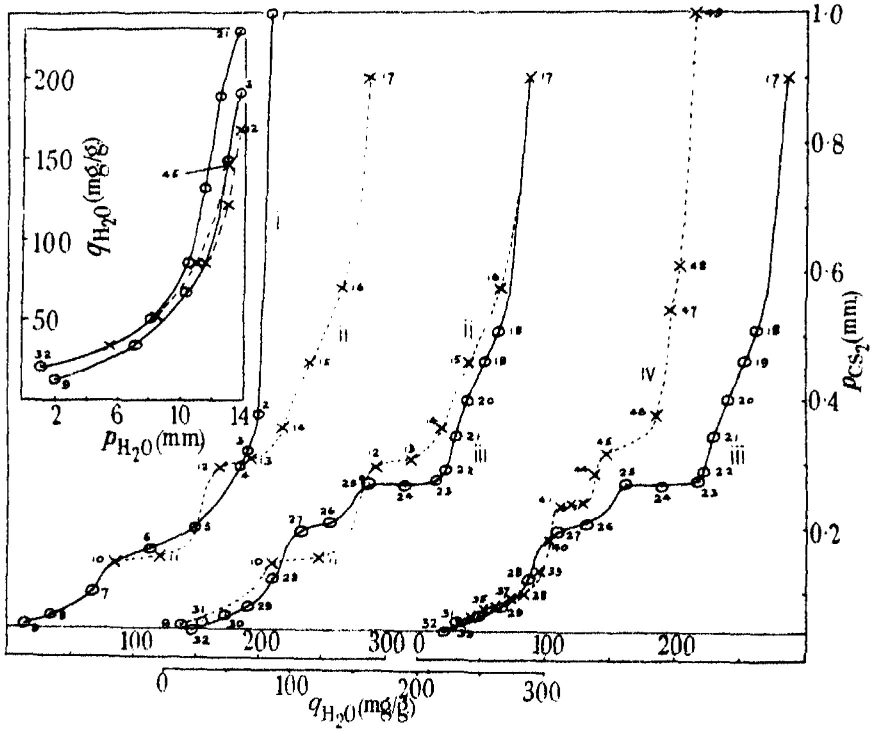


FIG. 5

Figures have been chosen in preference to tables for expressing the results, as no absolute value attaches to the data, and also because the discussion is easier to follow. Curve i in each of figs. 1–3 represents the isothermal for the water-free charcoal, derived from previous isostere measurements. Inset in fig. 4 are contained (*a*) the results found with low H_2O charges in Exps. 7–8, plotted on an expanded scale and (*b*) the aqueous vapour pressures observed during the same experiments, omitting the points where separation of liquid water in the neck of the container was observed. In fig. 5, containing the results of Exp. 9, the data for the first desorption (i) and the first resorption (ii), the first resorption and second desorption (iii), the second desorption and second resorption (iv), are plotted on three separate curves for an obvious reason. Inset are contained the corresponding data for aqueous vapour pressures, a few only of the actual points being indicated to avoid crowding the diagram.

Throughout, a number adjacent to an experimental point denotes the order of the reading inside the particular experiment.

DISCUSSION

It will be convenient to consider first the effect of changes in water content on the pressure exerted by a constant charge of adsorbed CS_2. *Figs.* 4 and 5 contain the data. Whilst the course of the curves in fig. 4 can hardly be doubted, in fig. 5, in spite of the large number of experimental points, there is admittedly an arbitrary element in the way the different curves are drawn. Nevertheless, one feature stands out plainly. The curves in fig. 4 and curve (i) in fig. 5 are composed of two, and the remainder of the curves in fig. 5 of three, loops. As the H_2O content of a charcoal is increased, the CS_2 pressure rises in stages, each characterized by an initial relatively low rate of pressure increase, which later becomes greater. Inset in figs. 4 and 5 are the corresponding water-vapour isothermals. They are identical in type with, and similar in position to, those given by the same charcoals in absence of CS_2 after degassing at temperatures of 270 and 800° (Allmand and others 1929*a*, *b*); isothermals on CS_2-free charcoals degassed at 110° are unfortunately not available. This close similarity suggests that the presence of CS_2 does not essentially affect the way in which the water is sorbed. This being so, it is natural to connect the loops in figs. 4 and 5 with the three stages, already mentioned, into which the sorption isothermals of H_2O on charcoals can be analysed.

Omitting complications due to the displaceable oxide layer, to the presence of hygroscopic ash, and to the existence of adsorption regions spatially difficult of access, we look on the taking up of increasing amounts of water vapour by a dry charcoal as follows. Water is first adsorbed essentially as single molecules, the pressure at the start therefore rising ever more and more rapidly, in the normal way, as the surface charge increases. Formation of isolated double molecules, exerting relatively less pressure, soon commences and, as they accumulate, so $\partial p/\partial q$ will become less, the isothermal showing a point of inflexion and becoming concave to the q axis. At a certain average water charge, local formation of a continuous water film will set in, this constituting the commencement of a second stage in water sorption, characterized initially by a still slower rate of rise of pressure, and only finishing when the whole of the available surface is covered with a condensed layer of associated water molecules. These two stages, whether in sorption, as described, or in desorption, can take place on a clean charcoal surface, so we think, at all events pseudo-reversibly, in the sense that a definite pressure, on either a sorption or a desorption isothermal, but not necessarily on both, will correspond to a definite total water charge, held in the ways described. Whether complete reversibility would result in complete absence of the complications mentioned is another matter. In any case, superposed on the second stage, and becoming, locally, physically possible when the surface is completely wetted, and thermodynamically possible when the conditions of the Kelvin equation are fulfilled, is capillary condensation. This, and the converse process, we imagine in practice to be largely irreversible, the filling and emptying of any particular set of capillaries depending not merely on the factors just mentioned, but on other and experimentally uncontrollable conditions, and taking place frequently at higher and at lower H_2O pressures respectively than would correspond to reversibility. The superposition of these phenomena on the continuous film formation already discussed means that, at higher q values, water isothermals will tend to be irreproducible in detail, and to contain sections where appreciable changes in q may take place practically independent of any change in p.

The pressure exerted by a definite charge of adsorbed CS_2 is a function of the extent and potential of the adsorbing surface which accommodates it, and this, in turn, will depend on the charge of sorbed water present, and on the way in which it is held. The results shown in figs. 4 and 5 are in general accordance with this view. The lower or lowest loop of the curves we regard as exhibiting the effect of variations in the charge of discontinuously adsorbed H_2O molecules on the fugacity of the adsorbed CS_2. We interpret the commencement of the second loop as indicating that formation of a continuous water film is commencing with, initially, as would be expected, relatively small effect on p_{CS_2}. Further increase in q_{H_2O} will result, sooner or later, in capillary condensation. This, in itself, will not affect p_{CS_2}. If therefore it occurs abruptly when any particular q_{H_2O} value is reached, a horizontal, or nearly so, section of the curve will result, i.e. a third loop will appear (fig. 5, ii, iii, iv). If, on the other hand, it accompanies the growth of the continuous water film, the slope of the curves will be less than in its absence, and, in general, any decrease in slope will indicate capillary condensation, as otherwise a progressive, if but slow, steepening would necessarily occur. Corresponding to what is known of H_2O isothermals on CS_2-free charcoals, it is to be expected that when, commencing with a high charge of H_2O, q_{H_2O} is diminished, desorption will take place first from the exposed continuous surface film, the capillaries emptying themselves, irreversibly, later, whereas when q_{H_2O} is increased, the capillaries will tend to fill relatively late. The result would be that p_{CS_2} curves for high values of q_{H_2O} would tend, during desorption of the latter, to be below the values found on sorption.

Fig. 4 contains two single water desorption curves, the results of Exps. 7 and 8. The CS_2 charges were almost identical in the two cases, the essential difference being that, in Exp. 7 the H_2O, and in Exp. 8 the CS_2, was sorbed first. The curves, besides being free from irregularities, are obviously very similar in nature. The shape of the H_2O isothermals for this charcoal (Allmand and others 1929*a*, *b*) suggests that, with the initial high charges of 256 and 306 mg./g. respectively, all the surface except that of lowest adsorption potential was coated with a continuous water film and that, in addition, there was considerable capillary condensation. The adsorbed CS_2 was therefore loosely held, exerting for example at a water charge of 250 mg./g. respectively about 77 times and 43 times its pressure on dry charcoal. Desorption results in removal of the surface film accompanied by an emptying of the capillaries, the former causing a rapid drop in p_{CS_2}. At 130 mg./g. the slopes of the curves become markedly less, only to steepen again from about 115 mg./g. onwards. We think that, in this region, the last patches of the continuous surface film of water disappear, but that before their removal is complete, a proportion of the residual adsorbed single and double water molecules has also been desorbed. The result is that the junction of the two loops of the curve, in the region of 100 mg./g. H_2O, is not well marked. Further loss of H_2O results in a progressive drop in p_{CS_2} to the "dry" charcoal value. As mentioned earlier, however, and as shown in inset (*a*), where the horizontal lines, as in the main diagram, denote these pressures for the original CS_2 charges, the very drastic treatment of the charcoal with water vapour has actually caused a small loss of adsorbed CS_2.

It will be observed (inset (*a*)) that, at the lowest q_{H_2O} values, the difference between the p_{CS_2} curves for the two experiments is no greater than would correspond to the difference in q_{CS_2}, i.e. the presence of the sorbed water has very little differential effect. In the same way, the H_2O isothermals in this region practically coincide, as is shown by the following figures (p. 40).

Exp. ...	8	7	8	7
q_{H_2O} (mg./g.)	4·6	7·1	11·8	14·8
p_{H_2O} (mm.)	0·003	0·006	0·139	0·195

But at higher H_2O charges, the H_2O isothermal lies at higher pressures in Exp. 8, where the CS_2 was sorbed first, than in Exp. 7 (see inset (*b*)); also above about 50 mg./g. of H_2O, the CS_2 pressures in Exp. 8 are lower than those in Exp. 7 to a far greater extent than can be accounted for by the smaller CS_2 charge. These results are all as might have been anticipated, in particular the effect due to the order in which CS_2 and H_2O were sorbed.

Exp. 9 (fig. 5) shows a more complex state of affairs, one however which can be briefly analysed. Curve i, allowing for the differences between charcoals *A* and *B*, as exhibited by their H_2O and CS_2 isotherms (the former with a more strongly adsorbing surface, and the latter with an exceptionally large volume of pores which can be filled with H_2O as a result of capillary condensation), closely resembles the curves in fig. 4, and can be interpreted in exactly the same way, the desorption of the continuous water film being completed when q_{H_2O} has fallen to about 85–90 mg./g. The sorption curve ii shows formation of the same film to commence at about the same q value. At about 165 mg./g., local and abrupt capillary condensation sets in, p_{CS_2} remaining practically constant, after which the curve rises continuously, but with a smaller slope than i, showing that considerable capillary condensation is accompanying the growth of the surface water film. The second desorption curve iii falls steeply at the start (delayed emptying of capillaries), and although it shows a significant inflexion of slope, it is not until it is appreciably lower than ii that an irreversible desorption of capillary bound water takes place (points 23–25).

The final flattening-out, marking the disappearance of the liquid film, is more pronounced than in i, and the last stage of H_2O desorption is reached both at higher q_{H_2O} (about 100 mg./g.) and p_{CS_2} values than on i and ii. Film formation on the second resorption curve (iv) commences at still higher values of p_{CS_2} and q_{H_2O} (about 105–110 mg./g.). Irreversible filling of capillaries takes place at some point near 150 mg./g., after which the curve rises more steeply than does ii, indicating a lesser degree of capillary condensation.

The movement of the lowest loop of the curves to higher q_{H_2O} and p_{CS_2} values during the course of the experiment we ascribe to the drift phenomena invariably noticed when studying water isothermals on charcoal. Between points 15–20, the charcoal stood for 10–11 days with a water charge of about 240–280 mg./g. Under these circumstances, not only will the charcoal surface have lost some water to the $ZnCl_2$ constituent, but also opportunity will have been given for more water to occupy relatively high-potential adsorption regions previously taken up by CS_2. Both these events will result in lower values of p_{H_2O} (fig. 5, inset). In addition, the former will move curve iii to the right as compared with ii, whilst the latter will explain the higher CS_2 pressures in the region of 90–100 mg./g. H_2O. The still higher CS_2 pressures observed at the end of the lowest loop in iv are probably due to further slow drift of the adsorbed water; three weeks elapsed between points 27 and 41.

Coming now to the CS_2 isotherms, fig. 1 contains the data for charcoal *A* with three different charges of water, as also the isothermal in absence of water (i). Remembering that the last was not directly determined, but "derived" from measurements of adsorption isosteres (Allmand and Lizius 1932) during which the surface oxide displacement was particularly effective, the isothermal ii given by the charcoal with 2 mg./g. H_2O coincides with that of the anhydrous charcoal within the experimental error, as would be expected from the experiments with the same charcoal already described. The small water charge, which exerts an exceedingly low pressure, is held on regions inaccessible to the larger CS_2 molecules. The isothermal iii on charcoal with 17·2 mg./g. H_2O (present as an unsaturated adsorbed layer—see fig. 4) lies at higher pressures throughout. At CS_2 charges exceeding 45–50 mg./g., it runs practically parallel to curve ii, but at lower pressures becomes considerably flatter than the latter. This is to be expected; the relative effect of a given H_2O charge is far greater at low, than at high, charges of CS_2. In addition, it will be noticed that on resorption the CS_2 pressures, even at high CS_2 charges, lie definitely above those on the desorption branch, a result clearly due to the occupation by the water of areas of high adsorption potential when the CS_2 charge was low (point 5).

Curve iv holds for charcoal containing 102·6 mg./g. of H_2O, a charge at which incipient continuous film formation is setting in (fig. 4). The effect on the pressure of small charges of CS_2 is consequently very great. Thus, p_{CS_2} at a charge of 8 mg./g. CS_2 is, initially, more than thirty times its value on water-free charcoal. And points 1–5 lie on a straight line (assuming point 4 to be in error) which is flatter than curve ii or the linear part of curve iii. At CS_2 charges exceeding about 65 mg./g., the isothermal commences to rise very rapidly indeed. This can only be due to the surface being very nearly completely covered at this stage with a continuous film of CS_2 and H_2O molecules. A comparison with the somewhat similar rapid rise of p_{CS_2} in the same pressure region observed in Exp. 7 (fig. 4), when q_{H_2O} exceeds 130 mg./g., shows that the effect of increasing the CS_2 charge in the present instance is far greater than that of increasing the H_2O charge in the experiment referred to. The following figures demonstrate the point:

p_{CS_2} mm.	q_{H_2O} ($q_{CS_2} = 54{\cdot}9$ mg./g.) mg./g.	q_{CS_2} ($q_{H_2O} = 102{\cdot}6$ mg./g.) mg./g.
0·33	130	60
0·50	143	70
0·79	159	76
1·26	193	80

The reason is, of course, that the whole of any increment in CS_2 affects the density of the surface charge, whilst much of the added H_2O is taken up by the capillaries and only indirectly affects p_{CS_2}.

As in the previous experiment, the values of p_{CS_2} at low CS_2 charges are higher on the second than on the first branch of the curve. The reason again can only be the sorption of H_2O on points previously occupied by CS_2. In this case, the time factor can have been the only one active; the unusual period of a week elapsed between the determination of points 7 and 8.

Fig. 2 presents isothermals for charcoal *B* containing 10·2 and 106·9 mg./g. H_2O, as also the "derived" isothermal for water-free charcoal. The effect of the smaller charge of H_2O is to displace the CS_2 isothermal throughout its course to higher pressures, more particularly, as before, when the CS_2 charge is low. Of considerable interest is the drift to lower pressures on desorption, clearly due to the slow uptake of H_2O by the hygroscopic $ZnCl_2$ present. The higher-charge of water, as in Exp. 3, corresponds to the initial stages of continuous film formation on the charcoal surface. The isothermal lies at still higher CS_2 pressures, as would be expected, and again shows a drift to lower pressures on desorption. The inflexions, which we have no reason to doubt experimentally, are difficult to understand.

Fig. 3 contains the results of a single experiment carried out on charcoal *C*, together with the isothermal for the anhydrous charcoal. In the absence of any parallel experiments done with a constant charge of CS_2, no detailed discussion is possible. The H_2O charge has a definite effect of the usual kind, relatively more marked at low CS_2 charges. The most interesting point is perhaps that its desorption and resorption branches almost coincide and, in any case, are considerably closer to one another than was found in the corresponding isothermal (fig. 1, ii) carried out on charcoal *A*. This we ascribe to the comparative absence (already mentioned) of the difficultly accessible adsorbing surfaces which are the cause of the "drift" effect observed when working with water.

Certain comparisons can be made between results given by identically charged charcoals in (*a*) experiments with constant H_2O charge and (*b*) experiments with constant CS_2 charge. The following are the data:

Charcoal	Exp.	Charge unchanged during exp.	H_2O charge in mg./g.	CS_2 charge in mg./g.	p_{CS_2} in mm.
A	2	H_2O	17·2	54·9	0·0708 (desorption)→0·0759 (resorption)
	7	CS_2	,,	,,	0·0413
	3	H_2O	102·6	54·9	0·275 (sorption and desorption)
	7	CS_2	,,	,,	0·294
B	4	H_2O	10·2	14·2	0·0569 (sorption)→0·0543 (desorption)
	9, i	CS_2	,,	,,	0·0475 (extrapolated)
	5	H_2O	106·9	14·2	0·194 (sorption)→0·188 (desorption)
	9, i, iii	CS_2	,,	,,	0·16–0·165 (i)→0·198 (iii)

In all cases, H_2O was put on the charcoal before any CS_2 was sorbed. Exp. 7 involved desorption of CS_2 only; to make the data more comparable, the figures for Exp. 9 are also taken from desorption curves.

The extent of agreement or disagreement between the different pairs of figures appears to be largely a question of the charge of water to which the charcoals were subjected during the different experiments. Where this was sufficiently high to displace, at all events, the bulk of the oxide complex from the charcoal surface, as was presumably the case in Exps. 3, 5, 7 and 9, agreement is quite as good as could be expected. Where however this was not so, as in Exps. 2 and 4, the CS_2 pressures are definitely higher than in the comparison experiments. This difference is particularly marked with charcoal *A*, a consequence of the highly developed oxygen charge on its surface, and the very high H_2O charge (256 mg./g.) initially present in Exp. 7. The difference between the results of Exps. 4 and 9, i with *B* may be partly due to the fact that an appreciable amount of the water in the latter was combined with the $ZnCl_2$ present.

SUMMARY

1. The simultaneous sorption of water and carbon bisulphide vapours by three different activated charcoals has been studied using an elaboration of technique already described.

2. Measurements have been made of the adsorption isothermals of carbon bisulphide on charcoals containing a constant charge of sorbed water, of the effect of variations in water charge on the pressure exerted by a constant

charge of sorbed carbon bisulphide and, over a limited range, of the water-vapour isothermals on such charcoals.

3. Minor complications were observed, due to slow displacement of surface oxides and of adsorbed carbon bisulphide by sorbed water, and to slow uptake of the latter by hygroscopic material present in one of the charcoals. The results in general are consistent with our previous knowledge of the different charcoals, and with the conception that water is sorbed on a clean charcoal surface at low pressures as a disperse layer of single and double molecules, that at higher pressures a continuous surface film is gradually formed which covers the whole available area at saturation, and that irreversible capillary condensation accompanies the formation of this film.

REFERENCES

Allmand, A. J. and Chaplin, R. 1930*a* *Proc. Roy. Soc.* A, **129**, 235–51.
— — 1930*b* *Proc. Roy. Soc.* A, **129**, 252–66.
— — 1931 *Proc. Roy. Soc.* A, **132**, 460–79.
— — 1932 *Trans. Faraday Soc.* **28**, 223–6.
Allmand, A. J., Hand, P. G. T., Manning, J. E. and Shiels, D. O. 1929*a* *J. Phys. Chem.* **33**, 1682–93.
Allmand, A. J., Hand, P. G. T. and Manning, J. E. 1929*b* *J. Phys. Chem.* **33**, 1694–1712.
Allmand, A. J. and King, R. B. 1930 *Proc. Roy. Soc.* A, **130**, 210–17.
Allmand, A. J. and Lizius, J. L. 1932 *Proc. Roy. Soc.* A, **134**, 554–70.
Allmand, A. J. and Puttick, A. 1930 *Proc. Roy. Soc.* A, **130**, 197–209.
Chaplin, R. 1928 *Proc. Roy. Soc.* A, **121**, 344–58.
— 1934 *Trans. Faraday Soc.* **30**, 249–60.
Hand, P. G. T. and Shiels, D. O. 1928 *J. Phys. Chem.* **32**, 441–55.

The band spectrum of silicon monosulphide and its relation to the band spectra of similar molecules

By R. F. Barrow and W. Jevons

Departments of Inorganic Chemistry and Physics,
Imperial College, London

(*Communicated by A. Fowler, F.R.S.—Received* 21 *July* 1938)

[Plates 1, 2]

Introduction

In a recent paper (Jevons, Bashford and Briscoe 1937) on the ultra-violet band system of GeO, a survey was made of the band systems of the monoxides and monosulphides of the Group-IV(B) elements which appeared to be analogous to the "Fourth Positive" system (A $^1\Pi \rightarrow$ X $^1\Sigma$) of CO and the well-known ultra-violet systems of CS and SiO. It was found that when the observed values of the vibrational coefficients ω_e and $x_e\omega_e$ were plotted against the number of electrons in the molecule two pairs of uninflected curves were obtained for the lower states of the monoxides and monosulphides, these states being probably all $^1\Sigma$ ground states, and two similar pairs of curves were obtained for the upper states. The only molecule in the subgroup for which data were entirely lacking was SiS; and for this a study of the curves suggested that there should be a system of ultra-violet bands degraded to the red, the heads of which would be represented roughly by

$$\nu_{\text{head}} = 37{,}000 + (550u' - 3u'^2) - (850u'' - 3u''^2),$$

where u' and u'' are written for $v' + \frac{1}{2}$ and $v'' + \frac{1}{2}$. Since each curve for the monosulphides has its maximum curvature between the points for CS and GeS, the estimates obtained by interpolation for the vibrational coefficients of SiS could not be more than a rough guide; and since the system origins show no smooth variation, but tend, as is clearly shown in a later section of the present paper, to follow the ionization potentials of the Group-IV(B) atoms, the prediction of the position of the $0 \rightarrow 0$ band (near $\lambda 2713$ according to the above expression) was also subject to some uncertainty. It was feared that the observation and measurement of band-heads of the predicted SiS system would be made difficult by the presence of overlying band systems, namely, the extensive S_2 system on the long-λ side, and, if a trace of oxygen or of carbon were present, the SiO system or the CS system on the short-λ side of the predicted system. Rather high temperatures would be required for the preparation and vaporization of either the monosulphide or the disulphide; the former was to be preferred for use in the discharge tube, since in the absence of excess of sulphur the S_2 emission bands might not be too strongly developed. An additional reason for the choice of the monosulphide was the hope of observing some of its bands in absorption, since the lower state in the predicted system should be the ground state ($^1\Sigma$) of SiS. In the SiS discharge employed, the CS and S_2 systems are, fortunately, entirely absent, and the SiO system, though present, has not been troublesome, because the strong bands of SiS happen to be about 150 A farther from the SiO than the above expression indicated, the $0 \rightarrow 0$ band being near $\lambda 2864$ instead of $\lambda 2713$ as roughly predicted. Some bands (with $v'' = 0$ and 1) in the ultra-violet system have been observed also in absorption by SiS vapour.

The present communication deals with the preparation of the monosulphide, the vibrational analysis of its ultra-violet system of which some 70–80 band-heads have been measured, and the observation and measurement of about the same number of bands in the near ultra-violet and visible regions. It will be followed by an investigation of the rotational structure of some bands of the ultra-violet system which have been photographed with high dispersion.

PREPARATION OF SILICON MONOSULPHIDE

The silicon monosulphide was prepared by the action of carbon disulphide on an excess of silicon heated to about 1000° C in a vitreosil tube S (fig. 1) 75 cm. long, which was surrounded centrally by an electric furnace F 20 cm. long, and connected into a closed pyrex circuit as shown. The carbon disulphide was allowed to distil, first in one direction and then in the other, over the silicon, unchanged carbon disulphide being condensed in a trap, T_1 or T_2, cooled in liquid air, the whole apparatus being exhausted through a third trap T_3 also cooled in liquid air. The wide pyrex end-pieces P_1 and P_2 were connected to the vitreosil tube with sealing wax, while apiezon M grease was found to be satisfactory for the conical joints C_1, C_2 and C_3 and the stop-cocks in spite of some action by the carbon disulphide vapour. The apparatus was so arranged that after preliminary evacuation and testing, the section C_1C_2 could be withdrawn as a whole for the introduction of the carbon disulphide into T_1 or T_2, and then replaced. Distillation was allowed to proceed until no carbon disulphide remained. The subsequent dismantling of the apparatus was, of course, facilitated by the removal of the section C_1C_2. The major part of the silicon monosulphide formed was deposited within 5 cm. from either end of the furnace, as a light yellow powder. At the centre of S a very hard black mass was obtained which contained much unchanged silicon and a small proportion of greyish white refractory crystals of silicon disulphide (Cambi 1910, 1911). Further, a small quantity of a dark red resinous substance, probably C_3S_2 (Stock and Prätorius 1912), was deposited in the traps T_1 and T_2 together with traces of a volatile material, which was probably carbon monosulphide. On the removal of the traps from liquid air, a mild explosion took place, which presumably indicated the spontaneous polymerization of the latter substance (Dewar and Jones 1910).

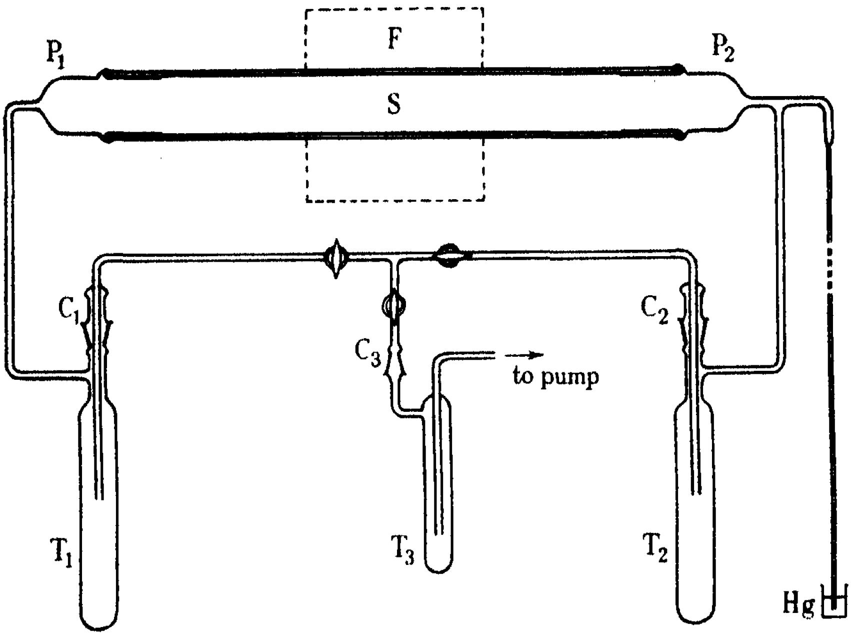


Fig. 1. Silica tube and pyrex circuit for preparation of silicon monosulphide.

The crude silicon monosulphide was purified by sublimation *in vacuo* in a silica tube heated to about 1000° C. The sublimate varied in appearance with the furnace temperature. At about 850° a small proportion (perhaps 10 %) of the material distilled from the tube and condensed as a pale yellow powder: at about 1050° the bulk of the material (90 % or more) distilled over and condensed wholly as a red vitreous mass: in distillations conducted at still higher temperatures, about 1150°, the condensate was almost black. When powdered, the dark-coloured glassy products become yellow or orange, and the experiments afford no evidence that silicon monosulphide ever exists as a black powder unless it is heavily contaminated with unchanged silicon. It is, therefore, a matter of doubt whether the two modifications and the transitions between them, described by Cambi, really exist. It may be that the so-called "black modification" is an association of silicon monosulphide with silicon, e.g. possibly a solid solution or a subsulphide; and it seems probable that the yellow powder and the red glass differ only physically in consequence of the differing rates of condensation. In any case the point is not material to the spectroscopic work, as all samples of the redistilled monosulphide when used in the discharge gave the same band spectrum with equal facility.

The silicon used in the preparation was examined spectroscopically in a carbon arc and found to contain a small quantity of iron, which, however, produced no observable effect in the subsequent spectroscopic work.

Quantitative chemical analyses were carried out on specimens of the monosulphide. Silicon was determined by burning a weighed quantity of the material in a heated platinum boat in a stream of oxygen. Sulphur was determined by breaking sealed bulbs containing known weights of the monosulphide in solutions containing a known quantity of iodine, and subsequently titrating the excess of iodine. In both determinations special precautions were taken to preclude action of atmospheric moisture on the sulphide. These determinations showed that the material contained $46{\cdot}5_5$ % of Si and $53{\cdot}4_4$ % of S, as compared with the calculated values for SiS, 46·67 and 53·33 % respectively.

APPARATUS AND OBSERVATIONS

In the first attempts to obtain an uncondensed discharge through the monosulphide vapour, an end-on vitreosil tube with a sealed-on window of clear silica was designed for insertion in an electric furnace at temperatures up to 1200° C. It failed because the sulphide attacked the heated electrodes (coaxial cylinders of sheet nickel) and also because there was much condensation of the monosulphide on the window. It was evidently desirable, therefore, to use a form of tube in which the electrodes and the window were much farther removed from the positive column of the discharge through the hot vapour. These conditions were fulfilled in the tube recently designed by Drs Pearse and Gaydon (1938), to whom the authors are much indebted for the suggestion that their device might be applicable to the present problem. As originally described and used by them, the tube consists of a horizontal pyrex envelope, about 80 cm. long and 3 cm. in diameter, with an end-on window, large aluminium electrodes some 30 cm. apart, and a narrow vitreosil positive-column tube supported axially in a steel carriage (a pair of centrally holed disks held about 6 cm. apart by a pair of screw-threaded rods) midway between the electrodes. For its original purpose, namely the production of spectra of hydrides of refractory metals, such as the spectrum of MnH, a small quantity of the powdered metal was introduced into a small central depression blown in one side of the vitreosil tube, and a stream of hydrogen at low pressure was passed through the tube. The current employed was such that the vitreosil tube became hot enough for the necessary volatilization of the metal, the whole pyrex envelope being immersed in a bath of running cold water.

For the present purpose argon was used instead of hydrogen as the residual gas, and cylinders of sheet nickel instead of aluminium for the electrodes, although one of the latter had an aluminium lining which acted as a "getter" for the nitrogen present in the argon. The powdered monosulphide was placed in the vitreosil positive-column tube. When the pressure of argon in the tube was suitably adjusted and a current of about 3 A passed through it the monosulphide began to distil into other parts of the tube, thus indicating that its concentration was sufficient for spectrographic observation. To photograph in a quartz spectrograph the stronger ultra-violet bands in the spectrum of the positive column, viewed end-on through a quartz window, very short times of exposure sufficed and the tube needed no further modification. The charge of monosulphide had, however, to be replenished after a few minutes run, and for this purpose the rubber stopper and one electrode were removed as a unit, the steel carriage and vitreosil tube were removed as another unit, and the apparatus reassembled after the recharging. With the tube in its original and very simple form, exposures with a grating, although only a few minutes in duration, had to be interrupted for the recharging. After the longer runs, also, it was found that small cracks occurred in the outer pyrex envelope near the points of support of the steel disks and near the ends of the electrodes. Two modifications were therefore made. To avoid cracking, the steel carriage and nickel electrodes were slightly reduced in diameter and placed in lengths of pyrex tubing, about 3 cm. longer than themselves, which served as linings for the pyrex envelope; one of these is shown at L in fig. 2. To avoid troublesome interruptions of a prolonged exposure for recharging, provision was made, as in fig. 2, for the occasional introduction of small amounts of SiS directly into the positive-column tube without "breaking vacuum". At the middle of the outer envelope *E* was a pyrex side tube *T*, about 1·5 cm. in diameter and 15 cm. long, with a greased conical ground joint *G* above the level of the water in the bath. To the upper tube at *G* was sealed, with wax, a pyrex tube *B* bent at right angles, the closed end of which contained powdered SiS, while the open end led into the widened upper end of a funnel *F* of vitreosil tubing. The lower end of *F* rested lightly on a short conical tube

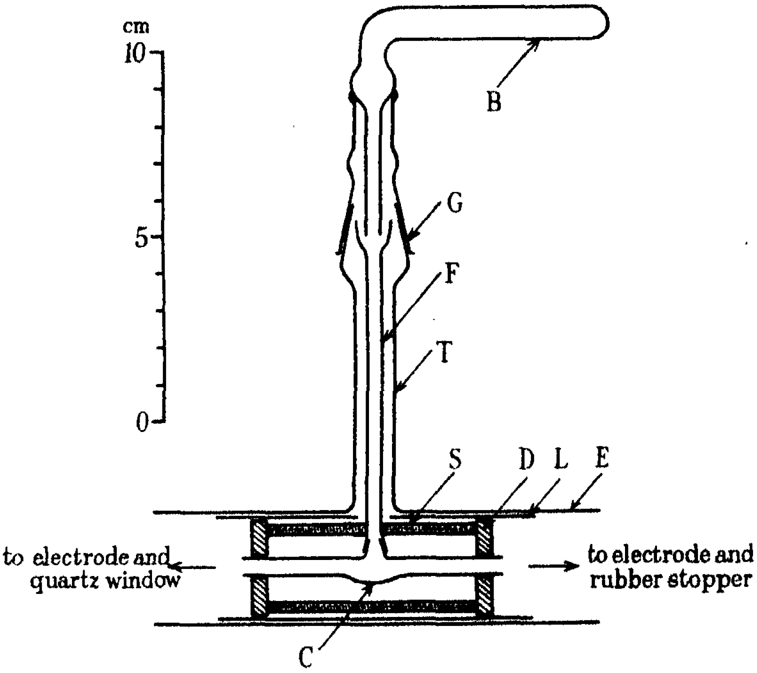

cm
10
5
0
B
G
F
T
S
D
L
E
to electrode and quartz window
to electrode and rubber stopper
C

Fig. 2. Section through middle of the modified form of Pearse and Gaydon's tube, *in a plane 45° from the vertical.*

which was sealed to the middle of the vitreosil column tube opposite the cup *C*. The envelope, which was placed horizontally in the water bath with both ends projecting through plasticine joints, was supported with the side tube *T* inclined at about 45° to the vertical, so that, when the tube *B* was suitably turned by means of the joint *G* and gently tapped, a little SiS fell through the funnel *F* into the cup *C*.

Current for the discharge tube was taken from a pair of 2400 V 5 kVA transformers whose primaries were connected in parallel to the 230 V supply of A.C. of 50 c./sec., and *whose secondaries were also connected in parallel.* Bands of the ultra-violet system were photographed against Fe arc lines on

Ilford Ordinary plates in (i) a large quartz Littrow spectrograph with a dispersion varying from 2·4 to 8·5 A/mm. in the range $\lambda 2570$–3870, and (ii) first and second orders of a 2·4 m. concave grating in an Eagle mounting with a first order dispersion of about 7·4 A/mm. With currents of about 3 A through the tube described above, the times of exposure were 2 min. upwards in the prism instrument and 5 min. upwards in the grating.

After the completion of the measurements and vibrational analysis recorded in the following section, work on the ultra-violet system was extended in several ways.

First, a short range of the ultra-violet absorption spectrum was investigated with the Littrow instrument by means of the continuum emitted by a heavy current (2 A) discharge in hydrogen in the tube described by Hunter and Pearse (1936). The absorption tube was of vitreosil, about 40 cm. long, of internal diameter about 2·5 cm. in the middle third and 1·0 cm. in the outer thirds, with waxed-on quartz windows. The wider part, into which a few grams of SiS had been introduced, was heated to about 1000° C in a nichrome-wired furnace, the tube being pumped out continuously before and during the exposures, which were of the order of 2 min. About a dozen band-heads having $v'' = 0$ and 1 were measured in absorption.* This observation not only supplies the desired proof that the lower electronic state is the ground state but is also a direct demonstration of the presence of unpolymerized SiS in the vaporized substance.

Next, with a view to an investigation of the rotational structure of selected bands of the ultra-violet system in emission, the type of discharge tube described above was abandoned in favour of an all-silica tube through which a heavy-current discharge could be maintained for much longer periods. This change was necessary because after prolonged running of the former tube the lower end of *F* (fig. 2) became closed by sublimed material. The silica discharge tube is of the end-on type shown in fig. 3. Its construction is facilitated by the avoidance of vitreosil-to-vitreosil joints, a piece of clear silica tubing, however short, being everywhere interposed between two pieces of vitreosil tubing. The large horizontal tubes containing the electrodes are of vitreosil about 30 cm. long and about 2·5 cm. internal diameter. One of them has a waxed-on quartz window *W*, and a branched side tube *T*, mostly of vitreosil, which serves the double purpose of admitting argon and supporting the pyrex sleeve S_1 which surrounds the nickel wire leading to the electrode E_1; the other is closed by a rubber stopper through which pass the pyrex sleeve S_2 of the other nickel lead and the exit tube leading to the pump. The electrodes are cylinders, about 15 cm. long, of sheet nickel lined with aluminium. The narrow central tube, in which the positive column of the discharge is concentrated, is of clear silica about 10 cm. long and about 0·6 cm. internal diameter, with a central enlargement to which is sealed a vertical vitreosil tube about 15 cm. long and 1 cm. internal diameter. To the upper end of this is connected (with sealing wax) a bent pyrex tube P with a ground conical joint G which supports a bent and closed tube containing the powdered SiS. With this arrangement SiS can be introduced at will into the positive column tube as before; in this case, however, the inlet tube is wide enough to remain unclogged by sublimation of the material. Currents up to 5 A may be passed through this tube for half-hour periods, and, with an occasional reduction of the current, the discharge has been run continuously for 2 hr. In this time several of the more important bands have been photographed on an Ilford Zenith plate in a fourth order of a 10 ft. concave grating in Eagle's original mounting, the dispersion being about 1·28 A/mm. The measurement and analysis of the rotational structure will be begun presently, but measurements of the band-heads on this plate are included in the present data (Table I).

---

* Professor G. Herzberg has kindly informed the authors that at the time of the appearance of the preliminary note (Barrow and Jevons 1938) he was about to start an investigation of SiS bands in absorption by large effective thicknesses of the vapour.

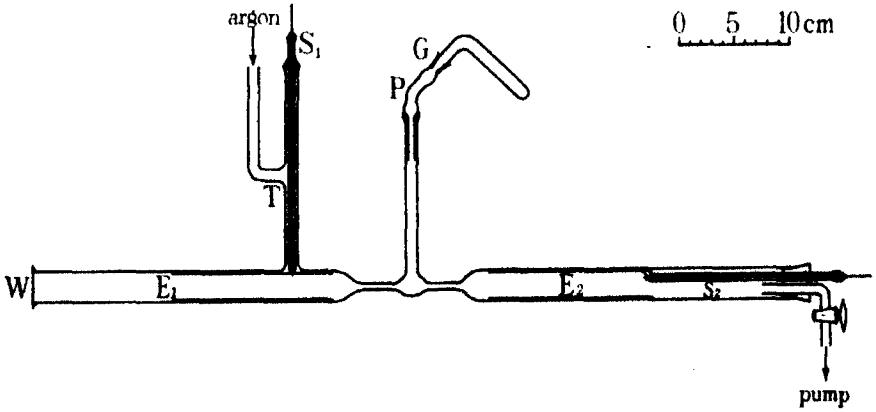


FIG. 3. All-silica discharge tube.

Finally, in view of the fact that silicon sulphide has been prepared by heating aluminium sulphide with silica (Tiede and Thimann 1926), it was not unreasonable to expect that the high temperatures (900–1000° C), to which the middle part of the silica discharge tube is heated by the large currents used, would suffice for the formation of SiS *in situ* and the development of the spectrum, if Al_2S_3 were used, instead of SiS itself, with argon in the tube. The experiment required only a slight modification of the tube, namely the addition of a silica lining to the central portion, with a hole in one side opposite the vertical side tube. With a relatively small current (about 0·5 A), and the tube fairly cool, the discharge spectrum consisted of the extensive $^3\Sigma \rightarrow {}^3\Sigma$ band-system of S_2 with little else but Al and Si lines, the latter being very weak. With a large current (4–5 A) through the hot tube, however, the ultra-violet system of SiS was strongly developed in addition to the S_2 system, and the lines of Si and Si^+ were also intensified; no bands attributable to AlS were detected. This result is of interest as showing that for the purpose of the present investigation the commercially obtainable Al_2S_3 might have been used instead of the specially prepared SiS, although the latter has the advantage that it gives no trace of the S_2 system. It is also of some importance as an indication that any attempt to produce a sulphide band spectrum in a high-current discharge in a silica tube may result in the development of SiS bands, which may thus prove to be a frequent "impurity" in such work,* just as are the CS bands in low-current discharges in the presence of C and S.

VIBRATIONAL ANALYSIS OF THE ULTRA-VIOLET SYSTEM

The most prominent feature in the ultra-violet spectrograms of the discharge is the SiS system of apparently single-headed bands degraded to the red, extending from about λ2585 to about λ3959. Partially overlapping the short-λ end of the system, and probably curtailing its measurement, are bands of the ultra-violet system $^1\Pi \rightarrow {}^1\Sigma$ of SiO, developed as a result of disintegration of the heated silica tube through which the positive column passes. There is no trace of the corresponding system of CS, the λ3064 system of OH, the extensive $^3\Sigma \rightarrow {}^3\Sigma$ system of S_2, or the analogous system of SO. Partially overlapping the long-λ end of the SiS system, in the region where it was feared that the S_2 system would be troublesome, are other hitherto unrecorded bands, which are described in the following section. Atomic lines present are due to Si, Si^+, Mg, Ni, Al, S, H and, of course, A. Littrow spectrograms of the SiS system are reproduced in fig. 4, Plate 1. Measurements of the band-heads, visual estimates of relative intensities and the assigned values of the vibrational quantum numbers are set out in a Deslandres scheme in Table I, where the bands observed in absorption as well as in emission are marked by *a*, and the high-dispersion (fourth order) measurements by *h*. It is possible to extend Table I in both directions (to $v' = 15$ and $v'' = 20$) to include a few other observed band-heads, which may belong to the same system but whose measurements were considered to be too uncertain to be recorded here. The wide Condon curve of intensity distribution is in satisfactory accordance with the rather large difference between the vibrational frequencies in the upper and lower electronic states. A study of the intervals in the upper state shows that the heads of the bands with $v' = 7$ are abnormally situated (about 10 cm.$^{-1}$ too low in wave-number), indicating perturbations of certain rotational levels associated with this vibrational level. If these bands are disregarded, the data can be represented approximately by an expression with a cubic term for the upper state only. The expression

---

* Further evidence of this has been obtained since the text of the paper was written. The SiS system is developed strongly in a discharge through a mixture of H_2S and BCl_3 vapour in an all-silica tube even with currents as small as 0·6 A, this being one of the sources used in an investigation of the band spectrum of BS now in progress.

$$\nu_{\text{head}} = 35028{\cdot}8 + (512{\cdot}0u' - 2{\cdot}38u'^2 - 0{\cdot}045u'^3) - (749{\cdot}5u'' - 2{\cdot}56u''^2),$$

where $u = v + \frac{1}{2}$, is an improvement on that given in the preliminary note on this work (Barrow and Jevons 1938). It has the merit that it represents almost perfectly the heads measured in the fourth order and associates the largest residuals ($\nu_{\text{obs.}} - \nu_{\text{calc.}}$) with a few outlying and relatively unimportant heads which may have been incorrectly located on account of superposed structure of neighbouring bands.

All the bands recorded are due to the most abundant molecule ^{28}Si^{32}S, none having been observed that could be ascribed to a less abundant isotopic molecule. This is not surprising in view of the relative rarity of ^{29}Si, ^{30}Si, ^{33}S and ^{34}S, as shown in the "International Table of Stable Isotopes" (1936).

NEAR ULTRA-VIOLET AND VISIBLE SYSTEM

On pp. 46 and 53 mention has been made of a set of emission bands occurring in the near ultra-violet and visible regions and partially overlapping the long-λ end of the ultra-violet system already described. They are well developed in the heavy-current discharge through SiS vapour and argon both in the modified Pearse-Gaydon tube (fig. 2) and in the all-silica tube (fig. 3), provided the SiS is present in excess. They were first observed in a visual instrument of small dispersion, in which they strongly resemble the S_2 bands of the extensive $^3\Sigma \rightarrow {}^3\Sigma$ system. They were afterwards photographed in a first order of the 2·4 m. grating, in which Ilford Special Rapid Panchromatic plates required exposures up to an hour in duration, the

TABLE I. DATA FOR BAND-HEADS OF ULTRA-VIOLET SYSTEM OF SiS

Italics in parentheses	Visual estimates of intensities.
Italics ...	Wave-lengths in air (I.A.).
Large Romans ...	Wave-numbers in vacuo (cm.⁻¹).
Small Romans ...	Wave-number differences (vibrational frequencies in upper and lower electronic states).

a	Measured in absorption as well as in emission.
h	Measured in high dispersion (fourth order of 10 ft. grating) in emission.
p	Perturbations: band-heads about 10 cm.⁻¹ too low.
n	Nebulous.
?	Observation doubtful or unusually difficult.
m	Band-head probably present but masked by background of rotational structure.

v'	v''→0		1		2		3		4		5		6		7		8		9		10		11		12		13		14		15		16		17		18
11					(2*n*) *2584·80* 38676·1	734·6	(2) *2634·85* 37941·5																														
					441·6		443·4																														
10					(2) *2614·66* 38234·5	736·4	(2) ? *2666·01* 37498·1																														
					451·1		449·0																														
9					(2) *2645·88* 37783·4	734·3	(2) *2698·32* 37049·1																														
					462·2																																
8			(2) ? *2626·88* 38056·6	735·4	(4) *2678·65* 37321·2																														(1) *3723·83* 26846·5	666·6	(2) *3818·64* 26179·9
			472·5																																467·1		468·1
7	(2) ? *p* *2608·29* 38327·9	743·8	(3) *p* *2659·91* 37584·1		*m*																								(2) *p* *3524·34* 28366·0				(2) *p* *3698·83* 27027·9	648·5	(2) *p* *3789·76* 26379·4	667·6	(1) *p* *3888·16* 25711·8
	461·8		462·6																										456·9				456·6				459·8
6	(4) *a* *2640·10* 37866·1	744·6	(2) *a* *2693·06* 37121·5																								(1) ? *3496·72* 28590·1	681·0	(2) *3582·04* 27909·1	671·2	(3) *3670·32* 27237·9	666·6	(2) *3762·40* 26571·3				(1) *3958·97* 25252·0
	477·0		477·7																								477·2		476·7		477·1		480·2				
5	(6) *a* *2673·78* 37389·1	745·3	(5) *a* *2728·17* 36643·8														(2) ? *3166·06* 31575·9						(2) *3391·40* 29477·9	683·5	(4) *3471·79* 28795·4	682·5	(3) *3556·08* 28112·9	680·5	(4) *3644·29* 27432·4	671·6	(2) *3735·75* 26760·8	669·7	(2) *3831·64* 26091·1				
	483·0		483·6																				478·5		487·1		486·4		479·6								
4	(7) *a* *2708·78* 36906·1	745·9	(6) *a* *2764·66* 36160·2												(2) *3143·52* 31802·3		*m*		(2) ? *3289·68* 30389·4	704·7	(3) *3367·78* 29684·7	685·3	(4) *3447·36* 28999·4	691·1	(3) *3531·53* 28308·3	681·8	(4) *3618·68* 27626·5	673·7	(2) *3709·14* 26952·8								
	490·6		489·5																488·0		483·8		487·8		489·5												
3	(8) *a* *2745·28* 36415·5	744·8	(2) ? *2802·60* 35670·7	740·5	(4) *2862·01* 34930·2						(3) *3053·03* 32744·9				*m*		(4) *3266·53* 30604·7	703·3	(5) *3343·37* 29901·4	700·5	(5) *3423·57* 29200·9	689·3	(4) *3506·35* 28511·6	692·8	(2) *3593·67* 27818·8												
	495·9		496·8		492·0												493·9		495·3		496·9																
2	(9) *a* *2783·18* 35919·6	745·7	(4) *a* *2842·18* 35173·9	735·7	(4) *2902·90* 34438·2	736·6	(5) *2966·35* 33701·6				*m*		(3) *3170·35* 31533·2	717·3	(5) *3244·16* 30815·9	705·1	(6) *3320·11* 30110·8	704·7	(5) *3399·68* 29406·1	702·1	(1) *3482·84* 28704·0																
	502·4		501·0				498·8						503·9		499·9		501·0																				
1	(9) *ah* *2822·66* 35417·2	744·3	(8) *h* *2883·25* 34672·9		*m*		(3) *3010·92* 33202·8	732·3	(7) *h* *3078·83* 32470·5	724·1	(7) *3149·05* 31746·4	717·1	(8) *3221·83* 31029·3	713·3	(7) *3297·64* 30316·0	706·2	(3) *3376·29* 29609·8																				
	507·1		507·2				510·6		507·1		505·8																										
0	(8) *ah* *2863·66* 34910·1	744·4	(10) *ah* *2926·06* 34165·7	739·3	(10) *h* *2990·78* 33426·4	734·2	(9) *h* *3057·95* 32692·2	728·8	(8) *h* *3127·67* 31963·4	722·5	(5) *3200·04* 31240·6																										

(Facing p. 54)

visible bands being much less intense than the bands near the middle of the ultra-violet system. They do not belong to the S_2 system or to any band-system known to the authors.

Parts of several of the spectrograms are reproduced in fig. 5, Plate 2. At the violet end the SiH bands near $\lambda 4142$ are strongly developed, and it would seem that the heavy-current discharge as now employed is a far more convenient source of these bands than the arc-in-hydrogen used by Jackson (1930). At the long-λ end the new bands are overrun by the H_2 bands and by atomic lines, especially those of A.

In the heavy-current discharge through Al_2S_3 in the silica tube, in which, as has already been mentioned (p. 52), the ultra-violet SiS system is intense, these less refrangible bands have not been detected at all, but the S_2 bands in this region are well developed, as also are the lines of Si and Si^+. Whether the new bands are entirely absent or are merely hidden by the stronger S_2 bands in this discharge it has not been possible to decide. Nor have they been detected in heavy-current discharges through H_2S, $H_2S + AlCl_3$ or $H_2S + BCl_3$ in the silica tube.

The band-head data are listed in Table II. Visual estimates of relative intensities are not included since the difference between the strongest and weakest bands appears to be comparatively little, the amount of overlapping is considerable, and the spectral region covered is large. The wave-numbers can be set out in a Deslandres scheme in which (i) the differences in the rows and columns appear to be in continuation of the vibrational frequencies for the high values of v' and v'' in the SiS ultra-violet system, the second differences being of the same order, namely about 5 cm.$^{-1}$, and (ii) the locus of the bands is a narrow diagonal limb of a Condon intensity distribution, which, at first sight, appeared to be a continuation of the long-λ limb of the ultra-violet system. This, however, has proved not to be the case, and the values of the vibrational quantum numbers v' and v'' are denoted by $p', p'+1, p'+2, \ldots$ and $p'', p''+1, p''+2, \ldots$, where p' and p'' remain undetermined. The assigned values of $v'-p'$ and $v''-p''$ are included in Table II and in fig. 5, Plate 2.

It is possible that these bands represent a second system of SiS, having neither state in common with the ultra-violet system. This may well be the case since SiS is expected to have not only singlet electronic states, two of which are involved in the ultra-violet system, but also triplet states, two of which may be involved in the system comprising these less refrangible "S_2-like" bands. It may be recalled that in the case of silicon oxide, too, a large number of near ultra-violet and visible bands have been recorded (Cameron 1927) in addition to the SiO ultra-violet system to which frequent

TABLE II. DATA FOR BAND-HEADS OF NEAR ULTRA-VIOLET AND VISIBLE SYSTEM

λ_{air}	$\nu_{\text{vac.}}$	$v'-p'$, $v''-p''$	λ_{air}	$\nu_{\text{vac.}}$	$v'-p'$, $v''-p''$
3491·26	28634·8	1, 0	4536·35	22038·0	6, 14
3571·65	27990·3	1, 1	4575·69	21848·5	7, 15
3623·41	27590·5	0, 1	4618·35	21646·7	8, 16
3654·71	27354·2	1, 2	4660·78	21449·6	9, 17
3687·02	27114·5	2, 3	4698·59	21277·0	7, 16
3709·14	26952·8	0, 2*	4740·66	21088·2	8, 17
3741·46	26719·0	1, 3	4748·77	21052·2	11, 19?
3774·30	26487·5	2, 4	4782·89	20902·0	9, 18
3806·35	26264·5	3, 5	4828·09	20706·4	10, 19
3831·64	26091·1	1, 4†	4873·77	20512·3	11, 20
3854·21	25938·3	7, 8?	4913·99	20344·4	9, 19
3864·50	25869·3	2, 5	4959·45	20157·9	10, 20
3885·43	25725·1	8, 9?	5004·64	19975·9	11, 21
3898·59	25643·0	3, 6	5051·98	19788·7	12, 22
3924·58	25473·2	1, 5	5095·09	19621·3	10, 21
3933·22	25417·3	4, 7	5143·35	19437·2	11, 22
3943·34	25352·1	7, 9?	5192·36	19253·7	12, 23
3958·97	25252·0	2, 6‡	5241·76	19072·3	13, 24
3978·99	25124·9	8, 10?	5287·68	18906·6	11, 23
3995·34	25022·1	3, 7	5336·46	18733·8	12, 24
4015·60	24896·0	9, 11?	5388·18	18554·0	13, 25
4029·57	24809·5	4, 8	5438·18	18383·4	11, 24
4064·65	24595·4	5, 9	5440·84	18374·4	14, 26
4075·30	24530·6	8, 11?	5484·82	18227·1	12, 25
4093·96	24419·4	3, 8	5490·24	18209·0	15, 27
4129·58	24208·7	4, 9§	5542·41	18037·7	13, 26
4164·14	24007·8	5, 10§	5596·96	17861·9	14, 27
4196·79	23821·0	3, 9§	5651·68	17689·0	15, 28
4234·00	23611·7	4, 10§	5703·26	17529·0	13, 27
4271·39	23405·0	5, 11	5759·25	17358·6	14, 28
4308·86	23201·5	6, 12	5811·60	17202·2	12, 27
4348·94	22987·7	7, 13	5816·45	17187·9	15, 29
4380·75	22820·8	5, 12	5876·12	17013·3	16, 30
4419·67	22619·8	6, 13	5988·35	16694·5	15, 30
4459·51	22417·7	7, 14	6052·38	16517·9	16, 31
4499·74	22217·3	8, 15	6112·85	16354·5	14, 30
			6169·34	16204·7	15, 31

* Also 4, 14 band of ultra-violet system.

† Also 5, 16 band of ultra-violet system.

‡ Also 6, 18 band of ultra-violet system.

§ Superposed on SiH band structure.

reference is made in this paper. They are far more complex than the bands described here and have not yet been interpreted; they may be due to a polyatomic emitter* such as SiO_2, while the present bands are almost certainly diatomic in origin.

ENERGY AND PRODUCTS OF DISSOCIATION OF SiS

In this and the following sections we return to the ultra-violet system. For the ground state, since the vibrational energies are fairly well represented by a quadratic in $(v''+\frac{1}{2})$, a rough estimate of the dissociation energy D'' may be obtained by linear extrapolation using the expression

$$D = (\omega_e - x_e\omega_e)^2/(4x_e\omega_e \times 8106)\,\text{eV}.$$

The result, namely $D''=6{\cdot}72$ eV, is probably too high an estimate as in many known cases, but falls well on a curve such as that given by Shaw (1937) for the monosulphides of this subgroup; for our present purpose we may perhaps assume $D''=6{\cdot}2\pm0{\cdot}5$ eV. For the upper state, since the introduction of an appreciable cubic term is necessary for even an approximate representation of the vibrational energies, a reliable estimate of D' cannot be obtained by extrapolation. All that can be asserted is that D' must be appreciably lower than the value, 3·37 eV, obtained by means of the above expression, i.e. by ignoring $y_e\omega_e(v+\frac{1}{2})^3$, and appreciably higher than the greatest observed vibrational energy, namely $G(15)$, which is found (by dividing the difference of the wave-numbers of the $15\rightarrow4$ and $0\rightarrow4$ band-heads by 8106) to be about 0·85 eV. Thus D' is well within the limits $2{\cdot}11\pm1{\cdot}26$, and may perhaps be taken as, say, $2{\cdot}1\pm0{\cdot}5$ eV.

We may now calculate the difference E_A of the total energies of the atomic products of dissociation of SiS in its excited and ground states. This difference, as is well known, is given by $\nu^{(0,0)}/8106+D'-D''$, which here amounts to $4{\cdot}32+(2{\cdot}1\pm0{\cdot}5)-(6{\cdot}2\pm0{\cdot}5)$, or $0{\cdot}2\pm1{\cdot}0$ eV. This would suggest that the actual energy difference may be zero; i.e. that the products of dissociation are the same for the upper state as for the ground state.

The ground state $(^1\Sigma)$ of CO, SC, SiO, GeS and PbO are generally regarded as dissociating into the ^{3}P ground states of C, Si, Ge and Pb and the ^{3}P ground states of O and S (see, for example, Sponer's "Molekülspektren, Tabellen", 1935). This is in accordance with Mulliken's (1932) general rule that the products of dissociation are the lowest-energy pair allowed by Hund's theory. The same will probably be true of the ground state of SiS;

---

* This seems to be suggested by improved photographs of one set of Cameron's bands, recently taken by R. C. Pankhurst in this laboratory.

and, if the above result be accepted, the upper state also of SiS will dissociate into the pair of ^{3}P ground states. Such a conclusion is in agreement with that already accepted for the A $^1\Pi$ state of CO and the corresponding $^1\Pi$ state of CS. It must, however, be examined more closely, in view of the wide limits of the above estimate of E_A, namely from $+1{\cdot}2$ to $-0{\cdot}8$ eV. We have to consider the possibilities that the upper state of SiS may dissociate into one atom in its ground state (^{3}P) and one atom in its first excited state (^{1}D), i.e. into Si(^{1}D) + S(^{3}P) with excitation energy about 0·75 eV, or into Si(^{3}P) + S(^{1}D) with excitation energy about 1·2 eV. Since these energies fall within the above limits, neither possibility is definitely excluded by the present results. However, if the upper state of SiS is, as seems almost certain, a singlet ($^1\Sigma$ or $^1\Pi$), both possibilities are excluded theoretically. For the requirement of conservation of resultant angular momentum of spin is fulfilled by the dissociation of a singlet molecular state into two triplet or two singlet atomic states, but forbids its dissociation into one triplet state and one singlet state. It seems probable, therefore, that SiS both in its ground and in its excited state dissociates into Si(^{3}P) + S(^{3}P).

DISCUSSION OF DATA FOR BAND-SYSTEMS OF RELATED MOLECULES

Data for the system origins and vibrational coefficients are now available for at least one band-system of each of the ten molecules of this subgroup of oxides and sulphides; and for each molecule there is a system which appears to have some analogy to the Fourth Positive system A $^1\Pi \rightarrow$ X $^1\Sigma$ of CO, this newly discovered SiS system now being added to the nine systems whose constants were studied graphically by Jevons, Bashford and Briscoe (1937). It seems desirable to extend these authors' study in the light of more recent information. First, rotational analyses of the SnO system A $\rightleftarrows$ X (Mahanti and Sen Gupta 1938) and the GeO system (Sen Gupta 1939) have shown the electronic transition to be $^1\Sigma \rightarrow {}^1\Sigma$ in each case, and not $^1\Pi \rightarrow {}^1\Sigma$ as has been already established for the CO, CS and SiO systems. Secondly, revised values of the coefficients have been derived (Jevons 1938) from Connelly's (1933) data and analysis of the A $\rightleftarrows$ X system of SnO; these coefficients for SnO and the coefficients given by Howell (1936) for PbO will be used here in preference to those used in figs. 2 and 3 of the former study (Jevons and others 1937), from which the constants for the remaining seven molecules (oxides and sulphides) are taken. Thirdly, Rosen and Désirant (1935) have published approximate values of the constants for CSe, for which they have observed an emission system similar to the $^1\Pi \rightarrow {}^1\Sigma$ system of CS. Fourthly, three absorptions systems of SnSe, three of PbSe and two of PbTe have been observed and analysed by Walker, Straley and Smith (1938). Thus, we now have data for the ground state and one or more excited states of thirteen of the group of fifteen oxides, sulphides and selenides, and one of the five tellurides. For reasons which will appear in what follows, it is probably the strongest system B ← X of SnSe, of PbSe and of PbTe that we should compare with the oxide and sulphide systems previously discussed. Finally, since some of the molecules of this group are isoelectronic with the Group-V(B) molecules for which spectra are known, the present discussion will gain in usefulness if both groups are included. In Table III are collected the constants for the ground states and the excited states under consideration. The symbols for the types of states are given only if they have been actually established by rotational analyses, although it would be reasonably safe to label all the ground states as $^1\Sigma$.

We examine first the electronic energies E_e of the excited states (i.e. $\nu_e/8106$ eV), which decrease irregularly as the electron content increases (fig. 3, 1937). In the case of the diatomic hydrides of the atoms of Group II (A and B), there is a surprisingly close proportionality of corresponding system-origins to the ionization potentials of the atoms—a fact for which, as Dr Pearse has remarked, no theoretical explanation appears to have yet been offered. The most direct way of applying a similar test to the present system is to divide the product of the ionization potentials, I_M and I_X, of each pair of atoms M and X, by the molecular excitation energy E_e and to observe the departure, if any, from constancy throughout the subgroup. The ionization potentials of the atoms concerned are given in Table IV; and the resulting values of $I_M\, I_X/E_e$ are shown in Table V, where figures for iso-electronic molecules appear in corresponding positions in (i) and (ii). It will be seen that they are of roughly the same order of magnitude throughout, but increase with mass and electron number of the Group-IV(B) atom both for the oxides and also for the sulphides (with the notable exception of CS), and decrease as we go from the oxide to the sulphide (again with the exception of CS) of the same atom. In view of the rather high value for CS and the surprisingly low ionization potential of Sn, it is safe to infer only that in this group the molecular electronic energies E_e have the same general trends as the ionization potentials though they are not strictly proportional thereto. The data are too meagre for any such statement to be made for Group-V(B).

We turn now to the vibrational coefficients ω_e and $x_e\omega_e$ for the ground and excited states. The two pairs of curves of ω_e against electron number (n) were found in the former graphical study (1937) to be of forms roughly resembling rectangular hyperbolae. Therefore, in order that graphs of

TABLE III. CONSTANTS FOR GROUND STATES AND COMPARABLE EXCITED STATES

Group IV(B) + VI(B) molecules

		Excited states				Ground states		
n	MX	State	ν_e	ω_e'	$x_e'\omega_e'$	State	ω_e''	$x_e''\omega_e''$
14	CO	A $^1\Pi$	65075·2	1515·6	17·25	$^1\Sigma$	2169·3	13·28
22	SiO	$^1\Pi$	42835·3	851·5	6·14	$^1\Sigma$	1242·0	6·05
40	GeO	$^1\Sigma$	37762·4	651·3	4·24	$^1\Sigma$	985·7	4·30
58	SnO	A $^1\Sigma$	29624·9	582·6	3·08	$^1\Sigma$	822·4	3·73
90	PbO	B $^1\Sigma$	22884·9	498·0	2·20	$^1\Sigma$	721·8	3·70
22	CS	$^1\Pi$	38912·1	1072·3	10·3	$^1\Sigma$	1285·1	6·50
30	SiS		36028·8	512·0	2·38		749·5	2·56
48	GeS		32889·5	375·0	1·51		575·8	1·80
66	SnS	B	28337·9	331·9	1·25	$^1\Sigma$	487·7	1·34
98	PbS	B	21847·8	282·2	0·856	$^1\Sigma$	428·1	1·20
40	CSe	? $^1\Pi$	(35250)	(830)		? $^1\Sigma$	1037	8
48	SiSe							
66	GeSe							
84	SnSe	B	22579·4	218·0	0·500		333·2	1·25
116	PbSe	B	21005·8	184·8	0·427		277·8	0·452
102	SnTe							
134	PbTe	B	19736·4	145·5	0·464		211·8	0·119

Group V(B) + V(B) molecules

		Excited states				Ground states		
n	MX	State	ν_e'	ω_e'	$x_e'\omega_e'$	State	ω_e''	$x_e''\omega_e''$
14	N_2	a $^1\Pi$	69623·1	1692·3	13·32	$^1\Sigma$	2359·6	14·44
22	PN	$^1\Pi$	39805·6	1103·1	7·22	$^1\Sigma$	1337·2	6·98
40	AsN	? $^1\Pi$	36009·1	863·0	8·24	? $^1\Sigma$	1068·0	5·36
58	SbN			(820)			(925)	
90	BiN							
22	NP	$^1\Pi$	39805·6	1103·1	7·22	$^1\Sigma$	1337·2	6·98
30	P_2					$^1\Sigma$	780·4	2·81
48	AsP							
66	SbP							
98	BiP							
40	NAs	? $^1\Pi$	36009·1	863·0	8·24	? $^1\Sigma$	1068·0	5·36
48	PAs							
66	As_2						429·4	1·12
84	SbAs							
116	BiAs							
102	Sb_2	A	33758	223·2	0·62		272·8	0·62
134	BiSb						220·0	0·50
166	Bi_2	A	17505	131·6	0·29		173·7	0·41

(i) The sources of data for Group V(B) molecules are: N_2 Lyman system, Jevons, "Report on Band Spectra" (1932); PN, Curry, Herzberg and Herzberg (1933); AsN, Spinks (1934); SbN, Coy and Sponer (1938); P_2, Herzberg (1932); As_2, Almy and Kinzer (1935); Sb_2, BiSb and Bi_2, Nakamura and Shidei (1935).

(ii) The BiN spectrum has been observed (Coy and Sponer 1938), but data are not yet available; SbN data are provisional.

(iii) The excited states of P_2 ($^1\Sigma$), As_2 (probably $^1\Sigma$) and BiSb (B, ? $^1\Sigma$), for which data are available, are apparently not analogous to those of N_2 (a $^1\Pi$), PN ($^1\Pi$) and AsN (? $^1\Pi$). See Jevons (1934). They are omitted from Tables III, V and VII.

TABLE IV. IONIZATION POTENTIALS (FROM H. BOMKE'S "VAKUUMSPEKTROSKOPIE", 1937) (eV)

Group IV(B)		Group V(B)		Group VI(B)	
$_6$C	11·203	$_7$N	14·460	$_8$O	13·549
$_{14}$Si	8·12	$_{15}$P	10·9	$_{16}$S	10·3
$_{32}$Ge	8·09	$_{33}$As	10·5	$_{34}$Se	9·5
$_{50}$Sn	7·30	$_{51}$Sb	8·4	$_{52}$Te	8·96
$_{82}$Pb	7·38	$_{83}$Bi	7·25		

TABLE V. VALUES OF $I_M I_X/E_e$

(i) Group IV(B) + VI(B)

	$_8$O	$_{16}$S	$_{34}$Se	$_{52}$Te
$_6$C	18·9*	24·0*	24·5†	
$_{14}$Si	20·8*	19·4		
$_{32}$Ge	23·5‡	20·5		
$_{50}$Sn	27·1‡	21·5	24·9	
$_{82}$Pb	35·4	28·2	27·1	27·2

(ii) Group V(B) + V(B)

	$_7$N	$_{15}$P	$_{33}$As	$_{51}$Sb	$_{83}$Bi
$_7$N	24·3*	32·1*	34·2†		
$_{15}$P	32·1*	§			
$_{33}$As	34·2†		§		
$_{51}$Sb				16·9	§
$_{83}$Bi				§	24·1

* Upper state $^1\Pi$.
‡ Upper state $^1\Sigma$.
† Upper state probably $^1\Pi$.
§ See note (iii) on Table III.

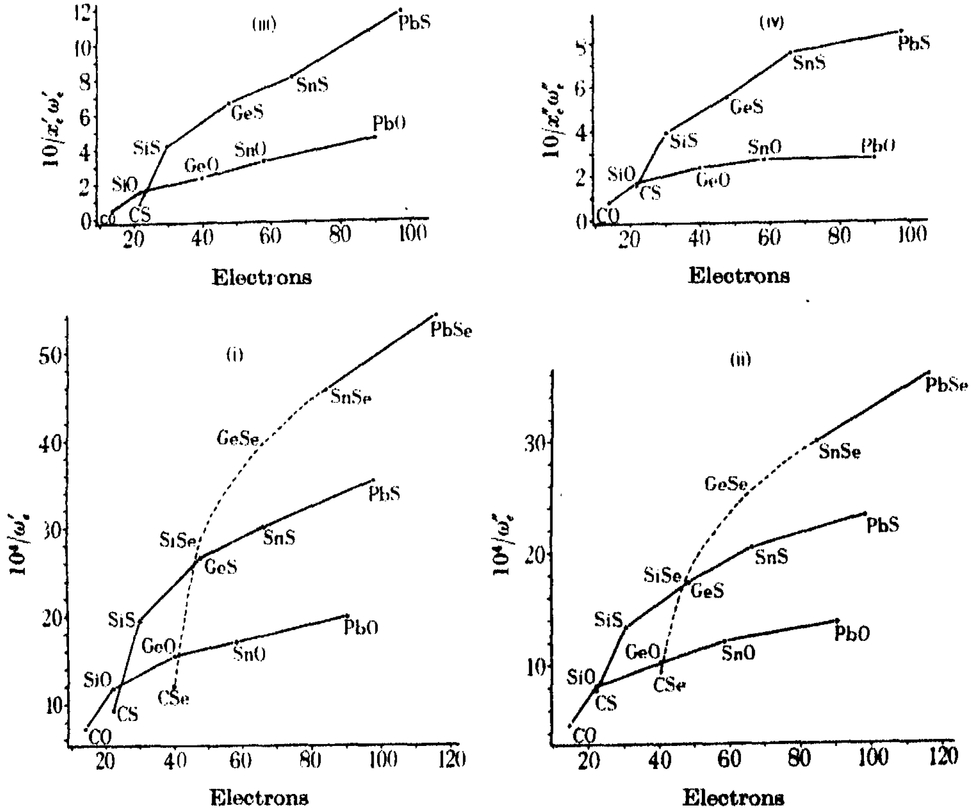
simpler form and smaller curvature might be obtained, the reciprocals of the coefficients, instead of the coefficients themselves, have now been plotted, as in fig. 6 (i) and (ii), where, however, straight lines are drawn between consecutive points, and the broken-line curves are drawn to indicate the values of ω_e to be expected for SiSe and GeSe. No such simple uninflected curves are obtained, however, by plotting $\log \omega_e$ against $\log n$.

Similarly, in the case of the anharmonic coefficients $x_e\omega_e$, if the reciprocals, instead of the coefficients themselves as in the former investigation, are plotted against n the resulting graphs for the oxides and sulphides appear to be of much the same simple form as the ω_e graphs. In view of the fact that a relatively small change of $x_e\omega_e$ produces an appreciable effect on the graph, even the approach to regularity shown in fig. 6 (iii) and (iv) is rather surprising, and it seems probable that the departures from uninflected curves can be ascribed to errors of observation.

If either the vibrational coefficients or their reciprocals are plotted against the reduced masses (μ) instead of the electron numbers, no simple curves are obtained.* Moreover, there is no smooth relation between ω_e and $1/\sqrt{\mu}$ or between $\log \omega_e$ and $\log \mu$. The latter tests were suggested by a recent observation by Carrelli and Trautteur (1937), that if the vibrational frequencies of the halogen molecules or of halides of metals in a given subgroup are plotted against $1/\sqrt{\mu}$, the points fall on a continuous curve which starts as a straight line through the origin and finishes as a parabola, the constants of the straight line and the parabola being different for the different subgroups.

---

* The apparent smoothness of the graphs for the sulphides previously noted, when reduced mass was used instead of electron number (Jevons and others 1937), is removed by the addition of the data for SiS.



Excited states

Ground states

FIG. 6. Vibrational coefficients of the ground states and comparable excited states of the sub-group of molecules.

Howell (1936) studied the vibrational coefficients ω_e for all the known electronic levels of PbO and PbS, and of SnO and SnS, in a way which may now be applied to the pairs of levels of all the molecules of this subgroup in Table III. First, Howell observed that the ratio of the coefficients ω'_e and ω''_e for an excited state and the ground state varies irregularly from one excited state to the next of the same molecule, but in the same manner for both molecules (oxide and sulphide) of the same element (Pb or Sn). In this way Howell recognized a probable correspondence between pairs of states of PbO and PbS, and also of SnO and SnS. In Table VI the values of this ratio for the recently recorded levels of SnSe, PbSe and PbTe are also included; Howell's rule can hardly be said to apply to these as convincingly as it does to the oxides and sulphides.

TABLE VI. VALUES OF RATIO ω_e'/ω_e'' FOR EXCITED STATES DENOTED BY ROMAN CAPITALS AND GROUND STATES X

SnO	SnS	SnSe	PbO	PbS	PbSe	PbTe
			F 0·517			
			D 0·734	E 0·699		
			? 0·643	D 0·664		
D 0·606	C 0·603	C 0·649	C 0·717	C 0·710	C 0·659	
A 0·708	B 0·681	B 0·657	B 0·690	B 0·659	B 0·665	B 0·687
B 0·755	A 0·779	A 0·672	A 0·626	A 0·610	A 0·601	A 0·668

The values of this ratio for the pairs of states of all the molecules in Table III are shown in Table VII, where again the figures for iso-electronic molecules appear in corresponding positions in (i) and (ii). In the present group (i), with the exception of the rather high figure for CS, they are close enough to suggest that we are in fact dealing with corresponding excited states, while in the Group V(B) molecules (ii) no regularity is discernible.

Howell's second observation is that the ratio of ω_e values for the ground states of PbO and PbS is very nearly the same as for those of SnO and SnS

TABLE VII. VALUES OF ω_e'/ω_e''

(i) Group IV(B) + VI(B)

	${}_{8}$O	${}_{16}$S	${}_{34}$Se	${}_{52}$Te
${}_{6}$C	0·699*	0·834*	(0·800)†	
${}_{14}$Si	0·686*	0·683		
${}_{32}$Ge	0·661‡	0·651		
${}_{50}$Sn	0·708‡	0·681	0·657	
${}_{82}$Pb	0·690	0·659	0·665	0·687

(ii) Group V(B) + V(B)

	${}_{7}$N	${}_{15}$P	${}_{33}$As	${}_{51}$Sb	${}_{83}$Bi
${}_{7}$N	0·717*	0·825*	0·808†	(0·886)	
${}_{15}$P	0·825*	§			
${}_{33}$As	0·808†		§		
${}_{51}$Sb	(0·886)			0·818	§
${}_{83}$Bi				§	0·758

* Upper state $^1\Pi$.
‡ Upper state $^1\Sigma$.
† Upper state probably $^1\Pi$.
§ See note (iii) on Table III.

and for those of GeO and GeS. This relation, which is a direct consequence of the foregoing, may be extended to the whole subgroup and should apply to corresponding excited states as well as to ground states. The result of the test is shown in Table VIII; the only abnormal values are those for the excited states of CS/CO and both states of CSe/CS; it will be seen that the mean values of the ratios increase as we go from MS/MO through MSe/MS to MTe/MSe. The same trend is suggested by the meagre data of Group V(B) as we go from MP/MN through MAs/MP to MSb/MAs.

TABLE VIII. RATIOS OF VIBRATIONAL COEFFICIENTS FOR MOLECULAR ELECTRONIC STATES IN TABLE III

	$\frac{\omega_e \text{ of MS}}{\omega_e \text{ of MO}}$		$\frac{\omega_e \text{ of MSe}}{\omega_e \text{ of MS}}$		$\frac{\omega_e \text{ of MTe}}{\omega_e \text{ of MSe}}$	
M	Excited states	Ground states	Excited states	Ground states	Excited states	Ground states
C	0·708	0·592	(0·774)	0·807		
Si	0·602	0·604				
Ge	0·576	0·584				
Sn	0·573	0·593	0·659	0·683		
Pb	0·568	0·592	0·655	0·649	0·787	0·762

The empirical relations recorded in this section are not put forward as evidence that what are described above as corresponding excited states are all of the same theoretical type, such as $^1\Sigma$ or $^1\Pi$. Indeed, rotational analyses of band structure, though carried out for relatively few of the molecules of these subgroups, have already shown that the excited states of CO, SiO, CS, N_2 and PN in question are $^1\Pi$, while those of GeO, SnO, PbO and P_2 are $^1\Sigma$. Such relations are useful, however, as a means of roughly predicting the constants for a molecule whose spectrum has yet to be observed. Thus Table V (i) suggests that for SiSe, $E_e \fallingdotseq 3{\cdot}86$ eV, $\nu_e \fallingdotseq 31300$ cm.$^{-1}$, and for GeSe, $E_e \fallingdotseq 3{\cdot}66$ eV, $\nu_e \fallingdotseq 29700$. From fig. 6 we may estimate the vibrational coefficients for these molecules: for SiSe, $\omega_e' \fallingdotseq 350$, $\omega_e'' = 550$, and for GeSe $\omega_e' \fallingdotseq 250$, $\omega_e'' \fallingdotseq 390$. That these estimates are reasonable is shown by a consideration of the resulting values of the ratio (ω_e of MSe)/(ω_e of MS), namely 0·68 and 0·78 for the upper and lower states, respectively, of SiSe/SiS and 0·67 and 0·68 for those of GeSe/GeS; these values are of the order to be expected from Table VIII. Again, the resulting values of ω_e'/ω_e'', namely 0·68 for SiSe and 0·69 for GeSe, are of the order required for Table VII (i). We may, then, with some confidence, expect SiSe and GeSe to yield

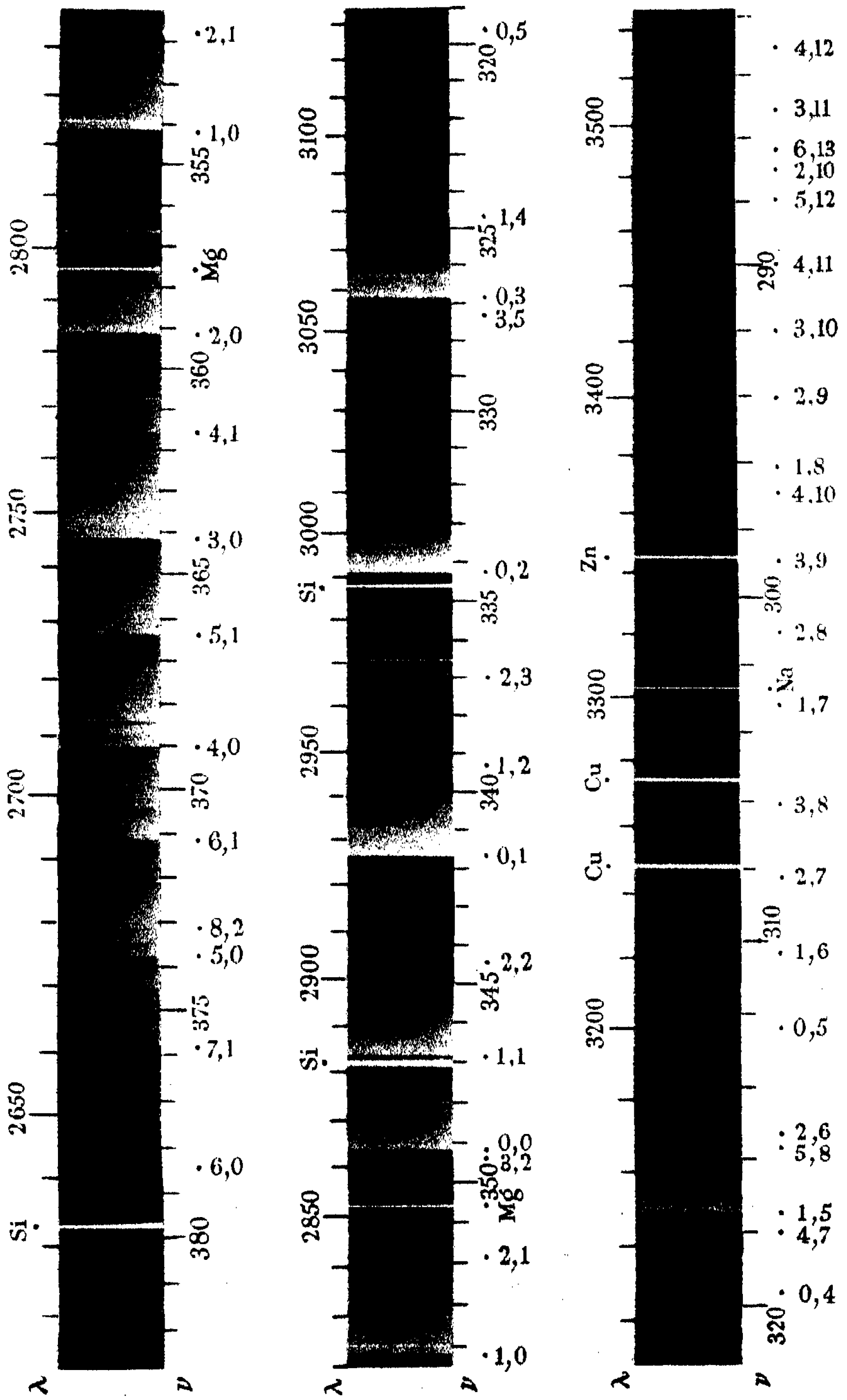


Fig. 4. Prism spectrograms of ultra-violet system of SiS. The numbers printed below each strip are values of v', v''.

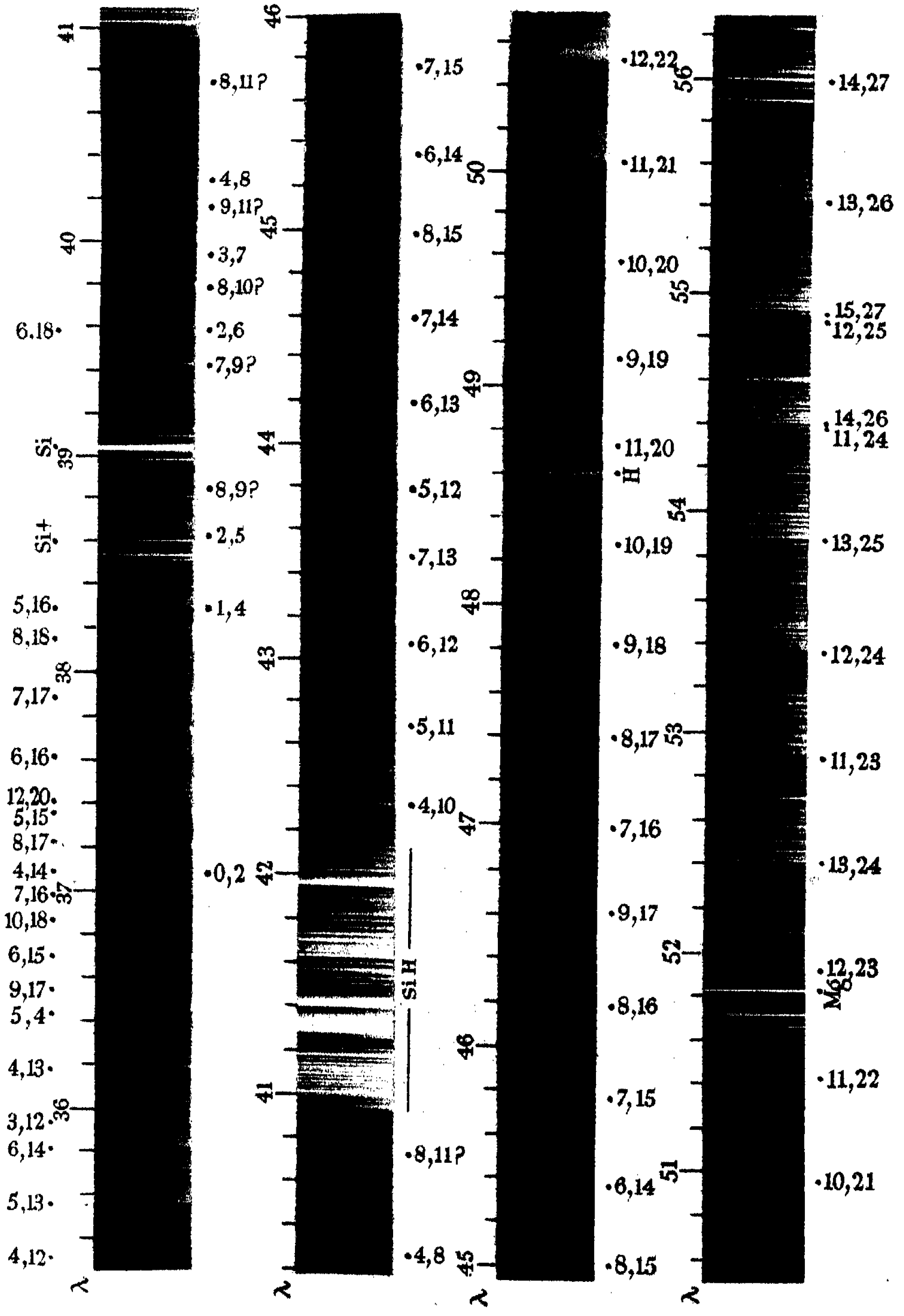


Fig. 5. 2·4 m.-grating spectrograms of near ultra-violet and visible system. The numbers printed below each strip are values of $v'-p'$, $v''-p''$ as in Table II. The numbers printed above the top strip are values of v', v'' for outlying bands of the ultra-violet system (fig. 4, Plate 1).

systems of ultra-violet bands degraded to the red, with heads roughly represented by:

$$\nu_{\mathrm{SiSe}} \fallingdotseq 31300 + (350u' - 1{\cdot}5u'^2) - (550u'' - 1{\cdot}7u''^2),$$

$$\nu_{\mathrm{GeSe}} \fallingdotseq 29700 + (250u' - 0{\cdot}8u'^2) - (390u'' - 1{\cdot}0u''^2),$$

the $0 \rightarrow 0$ bands being near $\lambda 3210$ and near $\lambda 3370$ respectively. Preparations are being made for searches for these systems.

The authors have pleasure in thanking Professor H. V. A. Briscoe for very valuable advice throughout the course of the work, and Professor H. Dingle for helpful discussion in the preparation of the paper. One of them (W. J.) is glad to express once again his indebtedness to the Fereday Electors of St John's College, Oxford, and the Government Grant Committee of the Royal Society.

References

Almy, G. M. and Kinzer, G. D. 1935 *Phys. Rev.* **47**, 721.
Barrow, R. F. and Jevons, W. 1938 *Nature, Lond.*, **141**, 833.
Bomke, H. 1937 "Vakuumspektroskopie." Leipzig.
Cambi, L. 1910 *R.C. Accad. Lincei*, **19** (ii), 294.
— 1911 *R.C. Accad. Lincei*, **20** (i), 433.
Cameron, W. H. B. 1927 *Phil. Mag.* **3**, 110.
Carrelli, A. and Trautteur, P. 1937 *Nuovo Cim.* **14**, 301.
Connelly, F. C. 1933 *Proc. Phys. Soc.* **45**, 780.
Coy, N. H. and Sponer, H. 1938 *Phys. Rev.* **53**, 495.
Curry, J., Herzberg, L. and Herzberg, G. 1933 *Z. Phys.* **86**, 348.
Dewar, J. and Jones, H. O. 1910 *Proc. Roy. Soc.* A, **83**, 526.
Herzberg, G. 1932 *Ann. Phys., Lpz.*, **15**, 677.
Howell, H. G. 1936 *Proc. Roy. Soc.* A, **153**, 683.
Hunter, A. and Pearse, R. W. B. 1936 *J. Sci. Instrum.* **13**, 403.
International Table of Stable Isotopes 1936 *Rev. Sci. Instrum.* **7**, 334.
Jackson, C. V. 1930 *Proc. Roy. Soc.* A, **126**, 373.
Jevons, W. 1932 *Phys. Soc.* 284. "Report on Band Spectra of Diatomic Molecules."
— 1934 *Nature, Lond.*, **133**, 619.
— 1938 *Proc. Phys. Soc.* **50**, 910.
Jevons, W., Bashford, L. A. and Briscoe, H. V. A. 1937 *Proc. Phys. Soc.* **49**, 543.
Mahanti, P. C. and Sen Gupta, A. K. 1938 *Z. Phys.* **109**, 39.
Mulliken, R. S. 1932 *Rev. Mod. Phys.* **4**, 1.
Nakamura, G. and Shidei, T. 1935 *Jap. J. Phys.* **10**, 11.
Pearse, R. W. B. and Gaydon, A. G. 1938 *Proc. Phys. Soc.* **50**, 201.
Rosen, B. and Désirant, M. 1935 *C.R. Acad. Sci., Paris*, **200**, 1659.
Sen Gupta, A. K. 1939 *Proc. Phys. Soc.* **51** (in the Press).
Shaw, R. W. 1937 *Phys. Rev.* **51**, 12.
Spinks, J. W. T. 1934 *Z. Phys.* **88**, 511.
Sponer, H. 1935 "Molekülspektren. I. Tabellen." Berlin.
Stock, A. and Prätorius, P. 1912 *Ber. dtsch. chem. Ges.* **45**, 3562.
Tiede, E. and Thimann, M. 1926 *Ber. dtsch. chem. Ges.* **59**, 1703.
Walker, J. W., Straley, J. W. and Smith, A. W. 1938 *Phys. Rev.* **53**, 140.

The behaviour of polar molecules in solid paraffin wax

By R. W. Sillars, D.Phil.

New College, Oxford

(*Communicated by E. V. Appleton, F.R.S.—Received* 25 *July* 1938)

Introduction

The idea that dielectric materials may contain minute electric dipoles, able to rotate under the action of an electric field, has been current for a very long time. Originally it provided simply a plausible explanation of the property which a dielectric has of increasing the capacity of a condenser into which it is introduced; later Hopkinson, Pellat, and others suggested that the "anomalous" after-effects of a condenser might be due to similar dipoles retarded by "frictional" forces and able to rotate only slowly towards alignment with the external field.

The quantity of electricity concerned in these effects is often approximately conserved, so that the explanation was an attractively simple one. With a growing knowledge of the time-scale of molecular processes, however, the notion that a molecular dipole might take seconds or minutes to take up an equilibrium position lost favour, while a quantitative theory of dipole effects was developed by Debye. From this it appeared that, for dilute solutions of polar substances in simple light liquids, the times involved would be of the order of 10^{-9} sec.

Measurements of dispersion and absorption in such liquids at frequencies up to the present limits of electrical measuring technique have largely confirmed Debye's quantitative conclusions. More viscous liquids, many of which contain molecules of widely differing sizes, and which may possess quasi-crystalline structures, do not show such a satisfactory agreement with the simple theory. More recently Debye and Ramm (1937) have published a treatment in which viscous and crystalline forces are regarded as acting simultaneously. This seems to fit the facts in some cases, but its application is somewhat limited because of the complexity of the analytical expressions involved.

It is now generally accepted that dipole rotation may take place in solids, and there is a growing body of evidence that the time constant involved may be comparatively large; anomalous dispersion and absorption attributed to this cause have been observed at audio-frequencies and even at

50 cycles/sec. These effects have been observed with such diverse substances as ice, glycerol at low temperature, *dl*-camphor, chlorinated diphenyls, and solutions of cetyl palmitate in paraffin wax. Although it may be debated whether these should all be regarded as solids in every sense, there is no doubt that the postulates of the original theory of Debye are not applicable to them.

In order to obtain more experimental information about effects of this sort, in the hope that it might help in formulating an adequate theory of them, it was decided to extend the experiments of Jackson (1935) with paraffin wax solutions. The long-chain paraffins are known to possess definite crystalline structures (Müller 1932) which are not likely to be seriously modified by the presence of a few per cent. of an ester of similar type. A preliminary experiment had suggested that the time constant depends to an enormous extent on the size of the ester molecule. The experiments of this paper were therefore directed to determining the time constants of 5 % solutions of different long-chain esters in a particular sample of paraffin wax at various temperatures below the solidification point.

For this purpose one may observe either changes of dielectric constant (*dispersion*), or changes of loss-angle (*energy absorption*). The former method is unsatisfactory with small concentrations of polar molecules, since the changes in capacitance due to changes in dielectric constant are of the same order of magnitude as those resulting from the changes of density and formation of voids in the wax on heating and cooling. The loss-angle method was therefore employed.

METHODS AND APPARATUS

The loss-angle of a condenser containing the material under test was determined at various temperatures and frequencies by the methods outlined below. The material was then removed and the condenser plates brought nearer together, so that the capacity was the same as before, and the loss-angle was again determined under the same conditions.

The difference between the two results, corrected for stray capacities in parallel with the specimen, was taken as the loss-angle of the material. The loss-angle of the empty condenser was found to be sensibly independent of temperature, so that it sufficed to measure this at room temperature only.

Radio-frequency measurements were carried out by the same method as Jackson employed. A suitable coil was connected to the terminals of the test condenser, forming a resonant circuit the power factor of which was found from the width of its frequency response curve (Moullin 1930, 1931).

Measurements at frequencies up to 25,000 cycles/sec. were effected with a bridge of the Schering type, using a Wagner earth.

The construction of the test condenser is shown in fig. 1. The sample is placed between two circular chromium plated brass electrodes. The upper electrode rests on the specimen unless the softening temperature of the latter is likely to be reached, in which case the electrode is supported from a brass block slung on quartz rods. The lower electrode is in good electrical and thermal contact with the copper case which surrounds the whole condenser. This case is surrounded by, and is in good electrical contact with, a tinned sheet bath, filled with oil, water or methylated spirit, which can be heated electrically, or cooled by the addition of ice or solid carbon dioxide. An outer tinned sheet case is packed with powdered cork, as lagging. This is connected to the copper case for radio-frequency measurements, and to earth for bridge measurements.

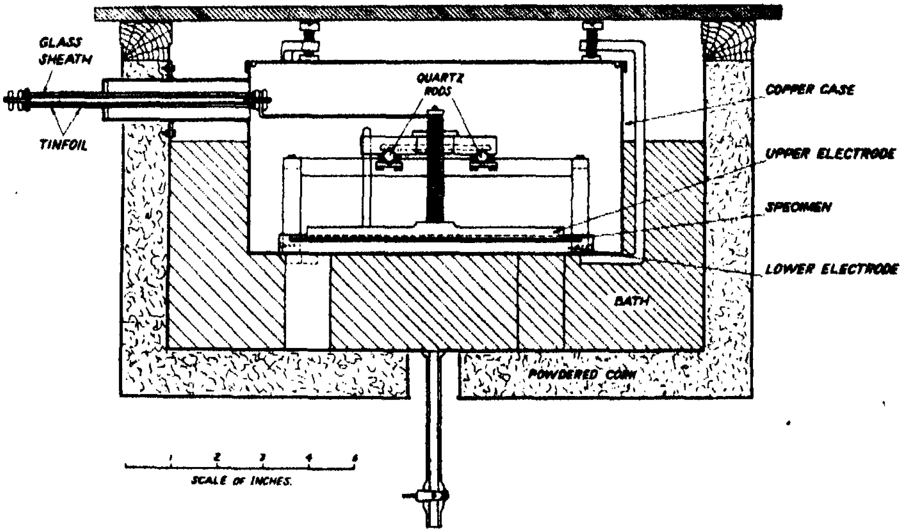


FIG. 1

Thermocouples are placed in the bath and in the upper electrode. The latter can be manipulated from outside the case and is withdrawn from the electrode before making a loss-angle measurement.

Connexion to the insulated electrode is made through a brass rod surrounded by a glass sheath protruding through the condenser case. It was found that when low temperature baths were used it was difficult to prevent water and snow from accumulating on the glass sheath where it emerged from the cold metal. As this lay in the stray field between the rod and the case, it gave rise to an indeterminate energy loss. To overcome this a strip of tinfoil was wound around the glass for about the first inch of its length, and connected to the case. This shielded the show from the stray field, and did not conduct away so appreciable a quantity of heat as to cause condensation farther out. Provided reasonable precautions were taken to wipe away moisture the leakage along the outside of the glass sheath was found to be insignificant, except for low audio-frequencies when it was eliminated by a further narrow strip of tinfoil connected to the earth of the bridge, using the familiar principle of the guard ring.

In order to prevent the entry of moisture into the condenser case all joints were packed with cork, and a supply of air, dried by passage through sulphuric acid and over phosphorous pentoxide, was pumped in through a side tube, a pressure of about 3 in. W.G. being maintained.

Presumably as a result of these precautions no trouble due to conduction by the surface of the specimen was encountered, and a guard ring was not employed. In one case a wax specimen which had been allowed to become damp showed a loss-angle at low frequencies roughly proportional to the inverse of frequency, but this appeared to be a volume rather than a surface effect, for the provision of a guard ring did not diminish it appreciably.

Both radio- and audio-frequency measurements were sensitive to about $0{\cdot}5 \times 10^{-4}$ in phase angle, and the probable relative error between points on the same curve was estimated to be about 10^{-4}. The absolute accuracy of the results depends on the accuracy of the determination of the loss-angle of the condenser, the accuracy of determination of stray capacities, and the absence of air gaps between the electrodes and the specimen. Errors due to the first two causes did not exceed 2 or 3 %, but it appeared that irregularity in the thickness of the specimen and more particularly the impossibility of ensuring that it lay perfectly flat, might give rise to errors as high as 7 % of the measured phase angle.

The difficulty of maintaining constant low temperatures in the bath gave rise to some uncertainty, particularly at temperatures far removed from that of the room. The procedure in taking readings was to observe the temperatures of the upper condenser plate and of the bath; when these were within half a degree of one another a measurement of loss-angle was made, and another pair of temperature readings taken immediately afterwards.

An uncertainty of about ± 1° C must be allowed in temperatures between 0 and − 30° C, and about ± 2° C in lower temperature readings.

MATERIALS

A refined wax melting over the range 57–60° C and known to have a negligible loss-angle was used as the solvent. Details of the esters employed are given below:

Ethyl stearate, $C_{20}H_{40}O_2$: commercial supply, used as received.

Butyl palmitate, $C_{20}H_{40}O_2$: synthesized from palmitic acid (Kahlbaum) which had been recrystallized from alcohol and benzene, and butyl alcohol of boiling point correct to $\frac{1}{10}$° C.

Butyl stearate (I), $C_{22}H_{44}O_2$: commercial supply, used as received.

Butyl stearate (II): commercial supply, twice redistilled.

Butyl stearate (III): synthesized from stearic acid (Kahlbaum) and butyl alcohol, and further purified.

Octyl palmitate, $C_{24}H_{48}O_2$: synthesized from palmitic acid and octyl alcohol.

Dodecyl palmitate, $C_{28}H_{56}O_2$: synthesized from palmitic acid and dodecyl alcohol.

Dodecyl stearate, $C_{30}H_{60}O_2$: synthesized from stearic acid (Kahlbaum) which had been recrystallized from alcohol and benzene, and dodecyl alcohol twice redistilled.

Cetyl palmitate, $C_{32}H_{64}O_2$: spermaceti recrystallized three times from acetone and three times from ether.

To prepare a specimen for loss measurements, the molten wax containing about 5 % of ester was poured on to a pool of mercury contained in an iron pot about 8 in. in diameter, which had been warmed to a temperature of 70 or 80° C. This was allowed to cool and the edges of the wax sheet so obtained were cut off leaving a sheet whose thickness was about 0·08 cm. and did not vary by more than $\pm$ 0·001 cm.

The sheets so obtained were not all similar in appearance, although their preparation was as far as possible identical. Those containing cetyl palmitate, dodecyl stearate, octyl palmitate, and butyl stearate (III) were uniformly translucent. The sheet containing butyl stearate (II) showed occasional patches of white opacity, those containing dodecyl palmitate, butyl stearate (I), and butyl palmitate had similar patches covering about half their areas, and the ethyl stearate specimen was almost entirely opaque.

It was suggested that these patches of opacity might be caused by the separation of some of the ester as a separate phase instead of remaining in solid solution. To investigate this point the solidification temperatures of solutions containing up to 20 % of butyl stearate (I) and of cetyl palmitate were observed. It was found that the depression of the solidifying point varied linearly with concentration, the slope of the temperature-molar-concentration graph being the same for both esters.

RESULTS AND DISCUSSION

Observations were made on all the esters at two or more frequencies and over a range of temperature sufficient to delineate the curve of tan δ down to values one-fifth of the maximum or less. A typical pair of curves, those for the specially synthesized butyl stearate, is shown in fig. 2. A set of similar curves covering five different frequencies is shown in fig. 3*a* for cetyl palmitate. The curves for the various esters were all fairly similar in shape with the exception of those of figs. 4 and 5 which will be discussed separately.

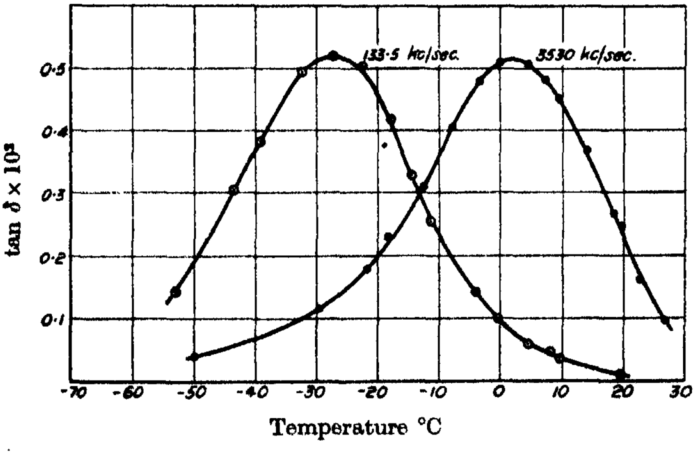


FIG. 2. 4·47% butyl stearate (III) in paraffin wax.

Observation of the variation of tan δ with frequency at constant temperature of 0° C were made on the cetyl palmitate specimen and are shown in fig. 3*b*. As it was more difficult to obtain accurate readings at constant temperature than at constant frequency, the constant temperature method was not employed again.

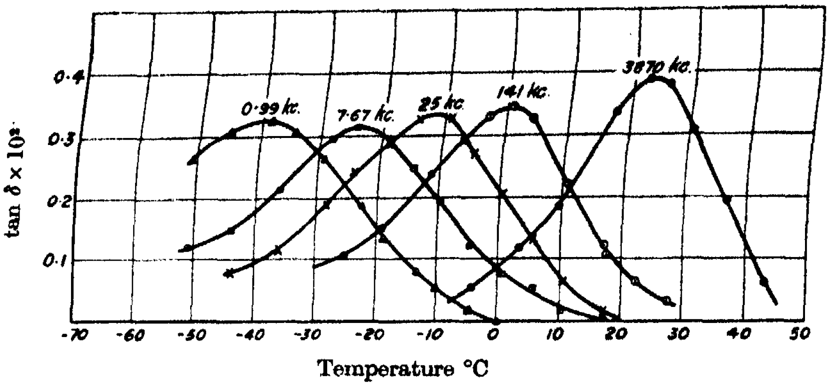


Fig. 3 *a*. 4·92% cetyl palmitate in paraffin wax.

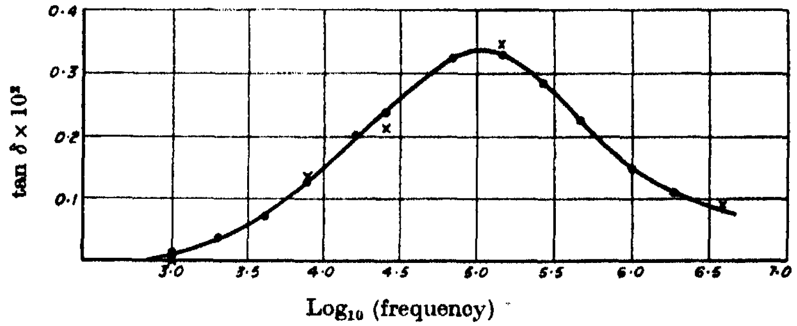


Fig. 3 *b*. 4·92% cetyl palmitate in paraffin wax at 0° C. × Points taken from constant frequency variable temperature curves.

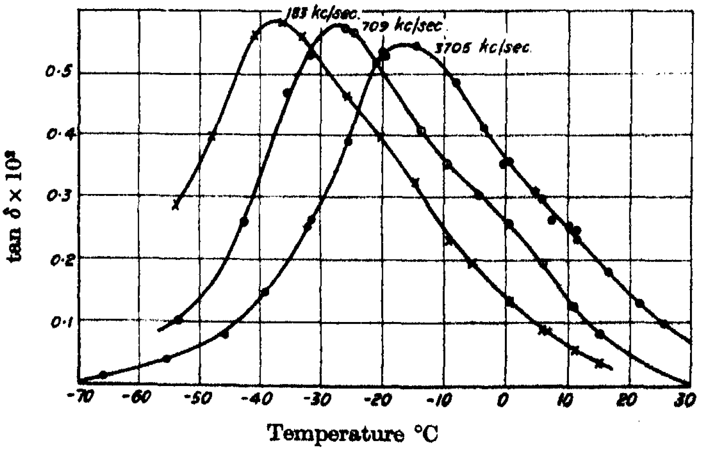


Fig. 4. 5·35% butyl stearate (I) in paraffin wax.

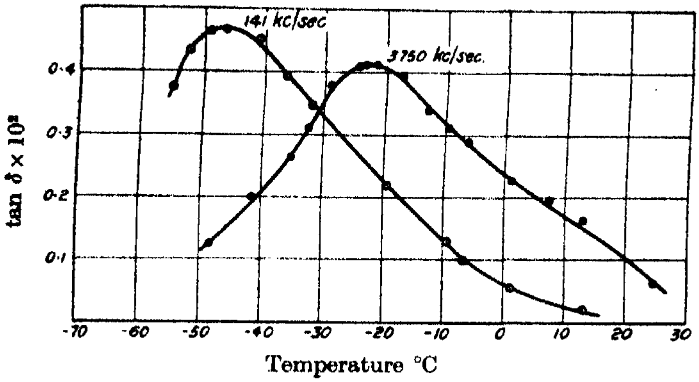


FIG. 5. 4·81% butyl stearate (II) in paraffin wax.

It is proposed to discuss these results in terms of Debye's equation

$$\kappa' \tan\delta = \kappa'' = \frac{4\pi}{3} \cdot \frac{(\kappa_0+2)^2}{3} \cdot \frac{n\mu^2}{3kT} \cdot \frac{\omega\tau}{1+\omega^2\tau^2}, \tag{1}$$

where κ' is the "real component of dielectric constant" (i.e. dielectric constant in the ordinary sense) at any frequency.

κ'' is the "imaginary component of dielectric constant" (corresponding to the conductivity or power loss),

δ is the amount by which the phase-angle between current and e.m.f. is less than 90°,

κ_0 is the dielectric constant at a frequency so low that dipole effects are fully developed,

n is the number of polar molecules per c.c.,

μ is their dipole moment,

ω is the angular frequency of the alternating e.m.f.,

k is Boltzmann's constant,

T is the absolute temperature,

τ is a constant with the dimension of time, corresponding to a relaxation time, henceforward the "time constant".

It will be assumed that the experimental curves are made up of a summation of curves of the type represented by equation (1). The use of this expression derived from consideration of rotation in a liquid finds more justification than appears at first sight, for any simple theory based on a relaxation mechanism leads to a similar form of equation. Debye, for instance, suggested the assumption that the ice molecule has two stable positions with a probability of transition from one to the other, and derived expressions of the same form as those of his liquid theory. The afore-mentioned treatment of quasi-crystalline liquids by Debye and Ramm results in a series of terms of the same type. Gemant (1935) has also published an analysis in which the dipole is supposed to be affected by elastic as well as viscous forces, and concludes that the form of the equation again remains the same as in Debye's original theory.

If for simplicity we write equation (1) as

$$\kappa'' = Cn\frac{\omega\tau}{1+\omega^2\tau^2}, \tag{2}$$

and assume that there are several groups of dipoles at work, the members of one group having time constants different from those of another group, but all having the same dipole moment, the total effect will be

$$\kappa'' = C\left(n_1\frac{\omega\tau_1}{1+\omega^2\tau_1^2}+n_2\frac{\omega\tau_2}{1+\omega^2\tau_2^2}+\dots\right).$$

In the more general case of a continuous distribution of the number of dipoles in any range of time constants such that

$$dn = F(\tau)\,d\tau,$$

we have

$$\kappa'' = C\int\frac{\omega\tau}{1+\omega^2\tau^2}dn$$
$$= C\int\frac{\omega\tau}{1+\omega^2\tau^2}F(\tau)\,d\tau.$$

If the curve of κ'' is plotted against $\log\omega$, the total area under the curve is

$$\int_0^\infty \kappa''.d(\log\omega) = \int_0^\infty\int\frac{\omega\tau}{1+\omega^2\tau^2}F(\tau)\,d\tau$$
$$= C\int\frac{\pi}{2}F(\tau)\,d\tau = C\frac{\pi}{2}\int dn$$
$$= \frac{\pi}{2}Cn. \tag{3a}$$

Thus this area is proportional to the total concentration of dipoles irrespective of the time constants associated with them. If the total concentration is known, then this area can be used to find the value of C, and hence the moment of the dipoles.

In the present experiments it is found that the curves are somewhat broader than equation (2) predicts, and this broadening is regarded as a spreading of the time constants about a central value. Such a spreading might be expected if the dipole molecules were not all identically situated. Whatever be the nature of the parameters controlling the value of τ, they seem likely, *a priori*, to be affected by imperfections in the crystal of the wax solvent. Such imperfections would occur in any case in a pure solid unless extraördinary precautions were taken; they will be much more common in the mixtures employed for the present experiments. It may be, indeed, that there exists no more than a certain orderliness in the arrangement of the molecules, not really meriting the description of a crystal. At any rate the existence of a range of time constants rather than a unique value is not surprising.

Equation ($3a$) provides a means of calculating the dipole moment irrespective of this spreading. A similar integration with respect to $\log\tau$ can only be carried out strictly if the form of $F(\tau)$ is known. It can be shown, however, that provided $F(\tau)$ approaches zero for all values of τ outside a certain small range, i.e. provided the range of spreading of τ is small compared with τ itself, its form is unimportant and

$$\int_0^\infty \kappa''.d(\log\tau) = \frac{\pi}{2}Cn \qquad (3b)$$

corresponding to ($3a$).

The quantity κ'' is equal to $\kappa'\tan\delta$. The dielectric constant of the wax alone was measured by Jackson (1935) and found to be 2·18. This value does not vary greatly with temperature below about 30° C, while variation due to dipole effects in the ester solutions is less than 1 %. In discussing the results of this paper, therefore, it will be assumed that

$$\kappa'' = 2{\cdot}18\tan\delta$$

with sufficient accuracy.

From equation (2) the maximum value of κ'' (or $\tan\delta$) for varying ω or τ occurs when $\omega\tau = 1$, so that at this point we have $\tau = \frac{1}{\omega} = \frac{1}{2\pi f}$. If, then, the temperature at which $\tan\delta$ is maximal is plotted against the logarithm of the inverse of the corresponding frequency, the resulting curve gives the relation between $\log 2\pi\tau$ and temperature. This is done in fig. 6, and a linear relation is found for dodecyl palmitate and for butyl stearate (I), while a slightly curved relation is found for cetyl palmitate. Most of the points of this figure are taken from curves of $\tan\delta$ against varying temperature. According to equation (1) varying the temperature also varies the constant

C of equation (2), so that maxima obtained in this way do not strictly correspond to $\omega\tau = 1$. For the same reason the values of $(\tan\delta)_{\text{max.}}$ should decrease with increasing temperature. Examination of the various curves shows no particular systematic variation of this sort however, nor was such an effect found by Jackson. This factor has therefore been neglected and it is assumed that variation of temperature varies τ only.

The relation between temperature and τ can also be obtained by comparison of a curve of $\tan\delta$ against temperature at constant frequency with one against log(frequency) at constant temperature. If at temperature T_0 $\tan\delta$ is maximal for $\omega = \omega_0$, and if a change of temperature to T_1 at constant frequency has the same effect as a change of ω to ω_1 at constant temperature, then the change of temperature T_0 to T_1 corresponds to a change of τ from $1/\omega_0$ to $1/\omega_1$. The points of fig. 7 are obtained in this way from figs. 3*a* and 3*b* for cetyl palmitate solution, the points of fig. 6 for the same ester are also shown as crosses on fig. 7 for comparison.

To determine the dipole moments from the curves of $\tan\delta$ against temperature using equation (3*b*), it was assumed that

$$\log\tau = sT + \text{constant},$$

where s is the slope of the appropriate line in fig. 6. The integral of equation (3*b*) is then the area under the curve of $\tan\delta$ against temperature, multiplied by s. The moments calculated in this way are tabulated below. A value for the moment of cetyl palmitate calculated from the area under the curve of fig. 3*b*, using equation (3*a*), is also included in the table.

Ester	Dipole moment	
Ethyl stearate	$1{\cdot}30 \times 10^{-8}$ e.s.u.	
	1·25 „	
Butyl palmitate	1·56 „	
Butyl stearate (III)	1·39 „	
	1·30 „	
Octyl palmitate	1·22 „	
	1·22 „	
Dodecyl palmitate	1·27 „	
Dodecyl stearate	1·39 „	
Cetyl palmitate	1·39 „	From $\tan\delta$-temperature curves
	1·38 „	
	1·33 „	
	1·36 „	From $\tan\delta$-frequency curves

The values obtained cannot be regarded as accurate, for in addition to the experimental errors already discussed the extrapolation of the curves to the

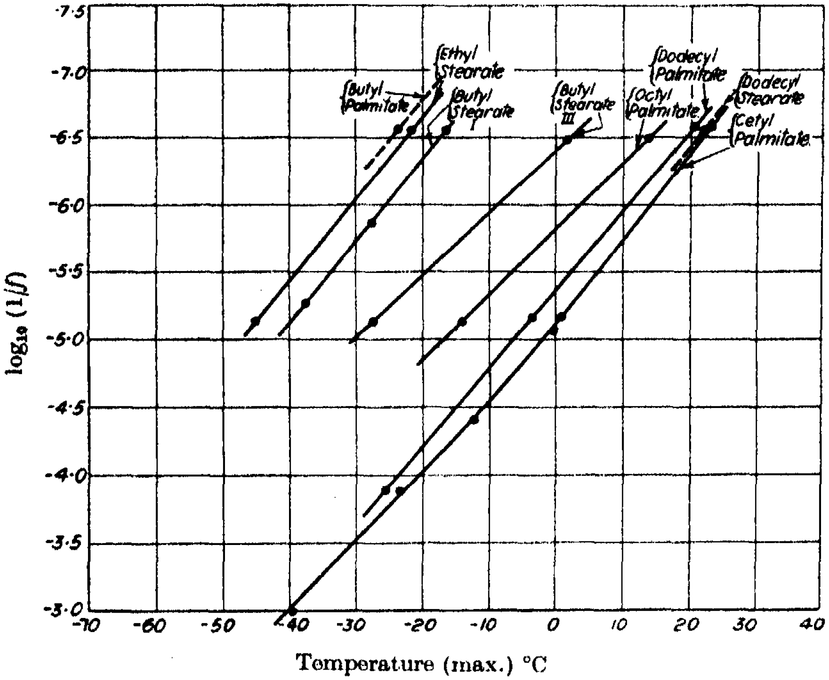


FIG. 6

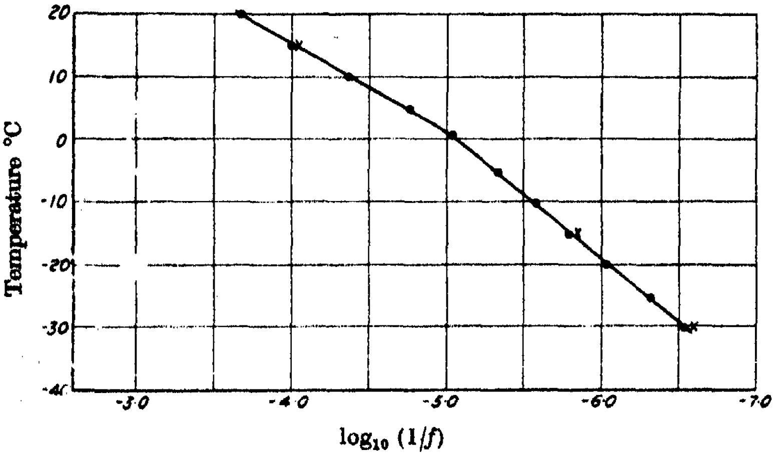


FIG. 7. Correspondence of temperature and log $2\pi\tau$ deduced from fig. 9 *b* and the 141 kc./sec. curve of fig. 9 *a*.

zero of tan δ involves a rather indeterminate possibility of error. They are noticeably lower than the moment of esters determined by the usual methods, which are in the neighbourhood of 1·6–1·8.

In the discussion which follows (p. 79) *it is concluded that rotation of the* dipole can only occur by rigid rotation of the whole ester molecule about its long axis. In this case one would expect only the component of the moment perpendicular to this long axis to be effective, and since it is considered on chemical grounds that the electric axis is not perpendicular to the geometric one, a somewhat low value of effective dipole moment is rather to be expected. Unfortunately there does not seem to be sufficient certainty regarding the precise angle between the electric and geometric axes to make possible an accurate comparison of the moments in the above table with the values found for esters in solution in liquids.

The value calculated by Jackson from his results with cetyl palmitate is higher than the corresponding figures of the table above, although he did not take into account the "spreading" effect. Jackson's curves are, in fact, higher and sharper than the majority of those of the present work. Similarly the 3870 kc. curve of fig. 3*a* is higher and sharper than those taken at lower frequencies. The cause of these differences has not been elucidated.

The majority of the curves are, however, of the general shape to be expected in view of the approximately linear relation between log τ and temperature, and allowing for the "spreading" effect already discussed. Exceptions are the curves of figs. 4 and 5, the former referring to a sample of butyl stearate as received from the supplier, the latter to a sample obtained from a different supplier and redistilled. The asymmetry of these curves might be attributed either to an abnormal and one-sided "spreading" effect, or to a non-linear relation between log τ and temperature*. The points for butyl stearate (I) in fig. 6 do not show any appreciable departure from linearity, so that this possibility is ruled out. It was ascertained from the supplier that sample (II) had been prepared from commercial stearic acid containing about equal parts of stearic and palmitic acids. Butyl palmitate and butyl stearate were therefore synthesized from the pure acids and alcohol, and the curves obtained from them confirmed the suspicion that the curves of figs. 4 and 5 have two components, one corresponding to butyl stearate, and the other, somewhat larger, to butyl palmitate which displays a larger apparent dipole moment. The effect of redistilling sample (II) was probably to increase the proportion of palmitate. In fig. 8 is shown a composite curve obtained by summing the effects of 2·88 % butyl palmitate and 1·92 % butyl stearate calculated by simple proportion from the appropriate curves; the similarity to fig. 5 is apparent.

---

* During some preliminary experiments on sample (I) the latter explanation was considered the more probable and a theoretical discussion by F. C. Frank (1936) gave further grounds for believing it to be correct. Subsequent investigation, however, led to rejection of his explanation of the gross asymmetry of these curves in favour of the one given here.

The elucidation of this point suggests that power factor measurements might provide a convenient means of identifying long-chain esters, ketones, etc. The separation and identification of mixtures of these compounds by ordinary chemical methods is extremely difficult (Piper and others 1931), but by dissolving the mixture in a hydrocarbon and plotting power factor against temperature or frequency it should be possible to identify each chain-length from its contribution to the curve.

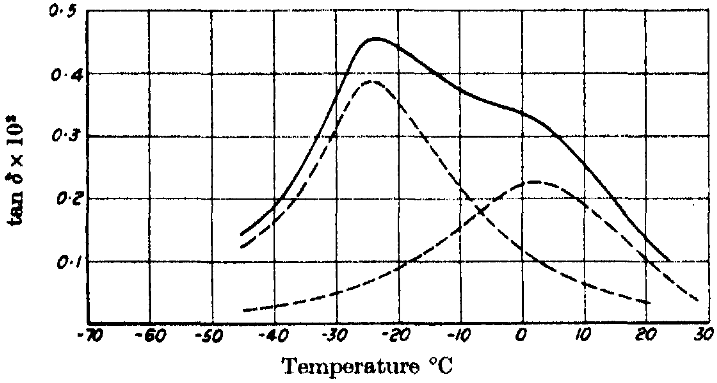


FIG. 8. 2·88 % butyl palmitate and 1·92 % butyl stearate.

It remains to enquire whether it is possible to elucidate the nature of the mechanism permitting rotation, and the physical significance of the time constant τ. It is known that pure hydrocarbon waxes form crystals in which the molecules lie side by side with their long axes parallel (Müller 1932). At low temperatures the packing is such that the molecules cannot rotate, but at temperatures near to the melting point the lattices of the higher waxes expand and rearrange themselves so that rotation of the molecules about their long axes can occur. It is considered that the molecule is a rigid structure and rotates as a whole. X-ray powder photographs were taken at room temperature of the wax used in these experiments, and also of the wax containing 5 % of butyl stearate (III). In both cases sharp rings were obtained indicating a definite crystalline structure. The side spacings seemed to indicate that the wax molecules themselves are unable to rotate, at any rate up to room temperature, and one must suppose that the rotating ester "fits loosely" into the lattice of fixed wax molecules. The proportion of ester molecules is not sufficiently great for a structural change in their immediate neighbourhood to show up on the X-ray photographs.

In fig. 9*b* the temperature for which $\tan\delta$ would be a maximum at a frequency of 3160 kc./sec. is plotted against the number of carbon atoms in the ester, the temperatures being obtained by interpolation in fig. 6. A smooth curve is obtained, from which it appears that the total length of the chain determines this temperature and that the position of the polar group in the chain is unimportant. Consider, for instance, butyl stearate having 22 carbon atoms. If we subtract two carbon atoms we can either take them from the long stearate chain leaving butyl palmitate, or from the short butyl group leaving ethyl stearate, and whichever we do the effect is the same, within the limits of experimental error. Similarly the point for dodecyl stearate in fig. 9*b* lies fairly between those for dodecyl and cetyl palmitates having two carbon atoms less and more respectively. It seems safe to conclude, therefore, that the molecule behaves as a rigid whole, for if rotation of the dipole were allowed by a distortion of the molecule one would expect the nearness or otherwise of the polar group to the end of the chain to have a considerable influence.

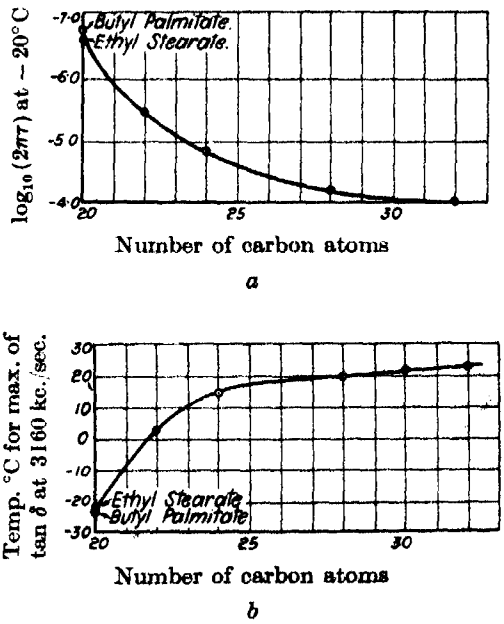


FIG. 9. Variation of time constant with chain length.

In connexion with fig. 9*b* it is perhaps worth remarking that the smooth curve provides additional confirmation that the effects observed are actually due to polar molecules and not to some spurious cause such as conduction effects arising from impurities. The processes of purification were by no means the same for all the esters and it is inconceivable that any spurious effects could give a smooth curve when plotted against the chain length of the ester.

As regards the significance of the time constant τ, the simplest hypothesis to make would be that the molecule suffers a "frictional" restraint, somewhat analogous to that of a shaft rotating in a journal. In this case a theory similar to that of Debye for liquids could be constructed, where the viscous restraint on a rotating sphere would be replaced by that on a rotating cylinder. The constant would then be proportional to the length of the molecule, i.e. to the number of carbon atoms in it. That this is not so is apparent from the table below. The values of τ shown are obtained by interpolation in fig. 6.

Ester	Number of carbon atoms	$2\pi\tau$ at $-20°$ C
Ethyl stearate	20	$2{\cdot}0 \times 10^{-7}$ sec.
Butyl stearate	22	$3{\cdot}3 \times 10^{-6}$,,
Octyl palmitate	24	$1{\cdot}6 \times 10^{-5}$,,
Dodecyl stearate	28	$6{\cdot}3 \times 10^{-5}$,,
Cetyl palmitate	32	$9{\cdot}8 \times 10^{-5}$,,

It is interesting to compare the results of Holzmüller (1937) who determined the time constants of a number of ketones ranging from acetone to nonyl acetyl ketone in solution in benzene and in hexane. He found that the time constants varied as a power of the molecular weight slightly greater than unity, and that the calculation of the time constant from Stokes' equation became more valid the higher the molecular weight. This, incidentally, also implies a rotation of the molecules as a whole, for if distortion occurred the time constant would increase less rapidly than the molecular weight.

Another way of interpreting τ is to regard it as inversely proportional to the probability that a dipole will rotate under the influence of the external field in a particular interval of time. This is the interpretation given to it by Debye in a tentative explanation of dipole effects in ice, to which reference has already been made. If, in the case of esters, we suppose each carbon atom to have an independent probability of moving, and if we further suppose that the dipole rotation can only occur if every atom of the molecule seizes an opportunity to move within a particular short interval of time, then the probability of rotation will vary exponentially with the inverse of the number of atoms, in other words log τ will vary as the number of carbon atoms in the ester. Fig. 9*a* shows log $2\pi\tau$ plotted against the number of carbon atoms, and though the departure from linearity is not so wild as in the attempt to correlate τ itself with the number of carbon atoms, there is no sign of support for the above supposition.

It seemed desirable to find out how the lengths of the wax molecules compared with those of the esters; to this end the long spacings were obtained from the X-ray photographs and found to indicate a chain length of about 26 carbon atoms. The influence of this magnitude on τ cannot be discussed in the absence of a reasonably detailed hypothesis regarding the mechanism of rotation. There is not, however, any satisfactory theory of dipole rotation in solids which takes into account the time effects which are the subject of these experiments, and it must be reluctantly confessed that these experiments have not suggested any simple extension of existing theory.

To Mr E. B. Moullin, for his invaluable help and criticism, and to Mr D. R. Pelmore, who carried out all the chemical operations, as well as the X-ray measurements, the writer is greatly indebted.

He also has to thank the Metropolitan-Vickers Electrical Company, Ltd., for a scholarship enabling this research to be carried out at the Engineering Laboratory of the University of Oxford, Dr A. P. M. Fleming, O.B.E., Director of the Company for his interest in the work, and the staff of the G.P.O. Research Station, Dollis Hill, for their help in connection with the X-ray measurements.

SUMMARY

The experiments described constitute an extension of previous work by Jackson and were carried out on solid samples of paraffin wax containing small concentrations of various long-chain esters. The esters are of molecular type similar to the wax, but have a permanent electric moment in virtue of the ester group. Dipole effects similar to those found by Jackson are observed and the variations of the characteristic time constants with temperature and with chain length of the esters are studied. Variation with temperature is approximately exponential, the exponent being about the same for each ester. It is found that the time constant is a non-linear function of the total chain length of the ester molecule, but is relatively independent of the position of the ester group in the chain. It is confirmed that the observed losses are due to dipole effects and not to some adventitious

cause. It is concluded that the whole ester molecule contributes to the time constant, and therefore that it behaves as a rigid structure and does not simply become deformed in the immediate neighbourhood of the dipole group. This is in agreement with evidence from other sources. The results do not suggest any simple relation between chain length and time constant, and a satisfactory theory to relate them has yet to be found.

References

Debye, P. and Ramm, W. 1937 *Ann. Phys., Lpz.*, **28**, 28.
Frank, F. C. 1936 *Trans. Faraday Soc.* **32**, 1634.
Gemant, A. 1935 *Phil. Mag.* **19**, 746.
Holzmüller, W. 1937 *Phys. Z.* **38**, 574.
Jackson, W. 1935 *Proc. Roy. Soc.* A, **150**, 197.
Moullin, E. B. 1930 *Wireless Eng.* **7**, 367.
— 1931 "Radio-Frequency Measurements", pp. 273 *et seq.* Griffin.
Müller, A. 1932 *Proc. Roy. Soc.* A, **138**, 514.
Piper, S. H. and others 1931 *Biochem. J.* **25**, 2072.

Ionic recombination in air

By J. Sayers, Ph.D.*

Cavendish Laboratory and St John's College, Cambridge

(*Communicated by E. V. Appleton, F.R.S.—Received* 26 *August* 1938)

1. Introduction

The fundamental importance of the problem of recombination of ions in gases was appreciated almost as soon as the role of ions as carriers of electric charge in gas conduction was recognized. Indeed it was at one time believed that the greater part of the light emitted in the Geissler tube discharge had its origin in the energy resulting from ionic recombination.

Rutherford and Thomson were the first to show, and verify experimentally, that the rate of decay of ion population by recombination was proportional to the product of the numbers of ions of either sign. Thus, if we denote by N_+ the number of positive ions per unit volume, and by N_- the number of negative ions, the recombination law established by them is

---

* Musgrave Research Student, Queen's University, Belfast.

$$\frac{dN_+}{dt} = \frac{dN_-}{dt} = -\alpha N_+ N_-,$$

where α, the factor of proportionality, is called the recombination coefficient.

Rutherford's methods, in somewhat modified form, were adopted later by other investigators with equal success. McClelland (1898) applied them in the determination of the rate of recombination of ions in gases drawn from flames and arcs; while later Townsend (1900) employed similar methods for measurements in a series of gases including air, hydrogen, oxygen, etc.

In more recent years there have been many determinations of α, but unfortunately there is but little agreement between the results obtained, this being particularly the case as regards the dependence of the rate of recombination on gas pressure. At high pressures of several atmospheres and upwards, it is generally agreed that the recombination coefficient is approximately inversely proportional to the gas pressure. But for pressures from 1 atm. downwards the evidence is very conflicting. Thirkill (1913) and Lenz (1932), for example, both find that α in air is a linear function of the pressure from 0 to 760 mm. Hg; while Schemel (1928) concludes from his experiments that α is almost independent of the pressure within this range.

The experiments here described were accordingly undertaken with the object of investigating the reasons for the lack of agreement rather than to add to the already confusing accumulation of data. With this end in view an apparatus was designed so as to permit various methods of procedure, in order that results by different methods might be compared under identical conditions.

In this paper attention is confined to cases where the negative ions are of atomic or molecular mass; and the chance of an electron recombining with a positive ion before attaching to a neutral molecule will be assumed negligible. It is easy to show that the probability of such an event is proportional to the quotient of ionization density divided by gas pressure, and for the range of these parameters here considered this can be entirely neglected.

2. EXPERIMENTAL METHODS

Two methods of measuring the recombination coefficient were investigated:

(i) The first was basically that designed by Rutherford (1897) and recently elaborated on by Marshall, Luhr, Mächler and others. In principle it consists in observing the rate of decrease of ion population in a gas which is ionized by a short flash of, say, X-rays or α-particles, and then left undisturbed.

In these experiments the ionization was produced by X-rays, and the gas was irradiated in short flashes controlled by a slotted brass disk which was made to rotate by a constant speed electric motor. To measure the ion concentration at any instant after the cessation of the X-ray flash, a high potential was applied to one of the electrodes of the ionization chamber and the ions collected and their charge measured. This collecting voltage was applied by a commutator device mounted on the same shaft as the rotating disk, and the phase relation of these two could be varied to give different intervals between the cessation of the X-rays and the application of the collecting voltage. The complete process was cyclic and repeated with each revolution of the disk and commutator. The pulses of charge received on the collecting plate each time the ions were swept out of the gas were smoothed out by a 0·0005 μF condenser and measured as a current. By multiplying this current by the period (about 0·7 sec.) of the cycle of events, the charge collected each time was estimated. From the rate of decay of ionization explored in this way it is easy to deduce the value of the recombination coefficient and the manner in which it varies with time.

(ii) The second method, due to Rümelin (1914), consists in exposing the gas continuously to the ionizing agency, and measuring the equilibrium value of ionization which is soon established between the rate of recombination plus diffusion, and the rate of production of ions. From a knowledge of the latter and a suitable allowance for the loss of ions by diffusion it is possible to evaluate the recombination coefficient.

3. SOME FEATURES OF THE APPARATUS

(i) The ionization chamber was constructed in accordance with modern high-vacuum technique for purification of the internal walls and metal electrodes. Instead of the more usual brass construction, it was built in a cylindrical envelope of pyrex, approximately 9 by 20 cm., the electrical connexions being brought out through vacuum-tight tungsten-pyrex seals. The metal disks forming the end-plates were of pure nickel and were held rigid and parallel by quartz spacers. The collecting plate was provided with a guard ring on which it was mounted with quartz supports, giving very high insulation resistance. A wire gauze was fitted round the inside of the pyrex envelope to prevent electrostatic charging of the glass surface giving residual electric fields inside the ionization chamber. This gauze was also made of pure nickel; thus all the metal parts inside the chamber were of the same material in order to avoid, as far as possible, the presence of contact potential differences.

After assembly of the electrodes, and just prior to sealing inside the pyrex envelope, they were carefully cleaned to remove surface contamination. When the tube was completed it was joined to the usual type of vacuum pumping train, consisting of mechanical pump, mercury diffusion pump, and liquid-air-cooled traps. The normal high-vacuum procedure of purification was then carried out, in which the entire tube was heated in an oven to 500° C for a few hours, and the metal parts were heated to redness by an induction heater. When gas was admitted to the tube and a test made, a small electric field was found to persist in the ionization space due to a residual contact e.m.f. This was due to two causes:

(*a*) The nickel gauze was apparently made of a nickel of slightly different purity from that used in the disks.

(*b*) Owing to the shielding effect of the guard ring, the inner collecting plate could not be heated to quite such a high temperature in the outgassing process as the other metal parts, so its surface was not quite so clean. As a final resort the small contact potential differences which still persisted, and were of the order of 0·1 V, were balanced out by small bias voltages introduced in the leads to the electrodes concerned.

(ii) The ionization current was measured by an electrometer type electron tube. It was used in conjunction with a galvanometer in a type of balanced bridge circuit due to Appleton (1928) and is illustrated in fig. 1.

The special advantage of this circuit is its great stability. When the resistance R_A in shunt with the galvanometer is made equal to the valve impedance and $R_1 = R_2$, the galvanometer does not respond to small variations of the voltage of the driving battery B_2.

Grid current in the valve was eliminated by proper choice of grid bias voltage. If, for example, the variation of grid current with grid potential, keeping the anode voltage fixed, is represented graphically, the curve is found to cut the axis at $-2{\cdot}0$ V. This value of grid bias was therefore chosen as the working point.

(iii) The electrical circuit of the ionization chamber and method of application of the collecting voltage have special features and so call for a brief description. The circuit is illustrated in fig. 2.

The first essential in this circuit is that no A.C. ripple from the application of the collecting voltage, or disturbance with higher frequency components from the commutator brushes, should reach the grid of the electrometer valve. Previous investigators have used electrometers for measuring the ionization current and these would not be much affected by A.C. disturbances; but in this case the rectifying properties of the valve make the above condition most important. The tube was accordingly made part of a balanced A.C. bridge circuit. With suitable choice of the condenser C and resistances X and Y it was arranged that the point A remained at earth potential as far as the application of the collecting voltage was concerned. The condensers

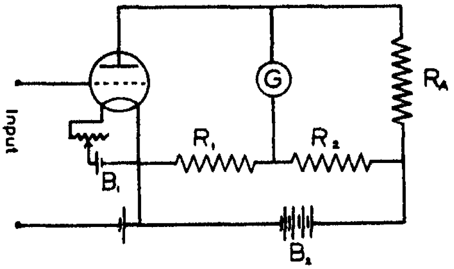


FIG. 1. Electrometer valve circuit.

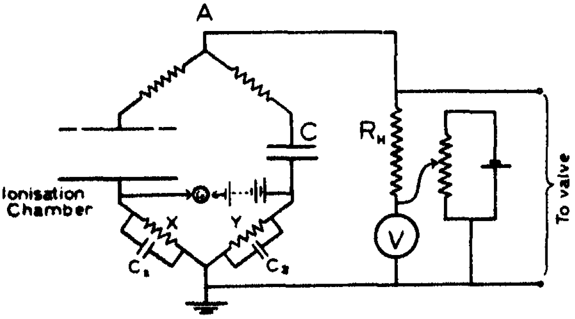


FIG. 2. Diagram of ionization chamber and associated electrical circuit.

C_2 and C_3 were found necessary to eliminate high-frequency disturbances caused by the make and break of the commutator. In the measurement of the ion current the valve was used as a null instrument. That is, the ion current was neutralized by a known and variable current supplied by a potentiometer and high resistance R_H, the valve circuit being used to indicate when exact balance was obtained.

4. THE DIFFUSION CORRECTION

Diffusion of ions to the collecting plates must always be considered as an effect which will make the measured value of the recombination coefficient too high. This is particularly so at low pressures, and precautions must be taken either: (i) To eliminate the escape of ions by diffusion, or (ii) to make due allowance for it in the calculations. As regards the first alternative, it may seem at first sight to be better, and might be achieved if it were possible to arrange that the region of gas ionized was always some distance from the plates; in this case no appreciable number of ions could diffuse to the plates in the time required to measure α. Unfortunately the volume ionized cannot be limited to a narrow section of the chamber by merely limiting the X-ray beam with lead slits. The reason for this is apparent when it is remembered that the greater part of X-ray ionization is a secondary process of ionization by β-particles, the latter being produced by X-ray photo-ionization in the body of the gas and also at the walls. These β-particles will produce a large amount of ionization exterior to the X-ray beam, and it is therefore inaccurate to assume that the volume ionized is just that defined by the lead slits. This mistake has been made by several investigators, and is an inherent error in the Langevin type of experiment for measuring recombination. The results obtained in such experiments are incorrect by a factor which is just the error in the volume assumed for the ionization. As a further check on this, experiments were performed with various widths of X-ray beam, and it was found that a more nearly consistent value for α was obtained by assuming that in all cases the ionization fills the chamber than by calculations based on an ionization restricted to the optical path of the X-rays.

It is therefore evident that the only method of dealing with the problem of diffusion is always to use the full volume of the ionization chamber, and calculate the diffusion correction to the observed value of α. The method adopted in these experiments is based on the following considerations:*

Consider a small element of volume of the ionization chamber between the planes x and $x+\delta x$.

Let n be the concentration of ions at the plane x.

Consider a small volume element of unit cross-section; the number entering the volume is

$$D\frac{dn}{dx},$$

where D is the coefficient of diffusion, while the number leaving

$$= D\frac{d}{dx}\left\{n+\frac{dn}{dx}\delta x\right\}$$

$$= D\left(\frac{dn}{dx}+\frac{d^2n}{dx^2}\delta x\right).$$

---

* Cf. J. J. Thomson (1928).

The nett loss of ions per sec. is therefore

$$= -D\frac{d^2n}{dx^2}\delta x. \tag{1}$$

Since everything may be assumed symmetrical about the central plane of the ionization chamber, we need only consider one half. Let the half depth be l. Then we have at central plane $x=0$, and at the plate $x=l$. To find the magnitude of the diffusion loss assume that it alone is responsible for the loss of ions, and let there be an equilibrium set up between this and a rate of production of q ion pairs per c.c. per sec.

Then $$q+D\frac{d^2n}{dx^2} = 0. \tag{2}$$

Integrating this equation, and remembering that

$$\frac{dn}{dx} = 0 \text{ at } x = 0,$$

because of symmetry, we get $$\frac{dn}{dx} = -\frac{q}{D}x. \tag{3}$$

The solution of this, subject to the boundary condition $n=0$ for $x=l$, is

$$n = \frac{1}{2}\frac{q}{D}(l^2-x^2). \tag{4}$$

Now, the total number of ions between the plates is

$$N = 2A\int_0^l n\,dx,$$

where A = area of plates, i.e.

$$N = 2A\int_0^l \frac{1}{2}\frac{q}{D}(l^2-x^2)\,dx$$

$$= \frac{Aq}{D}\frac{2}{3}l^3, \tag{5}$$

or the average number per c.c. is

$$\frac{N}{Al} = \frac{2}{3}\frac{ql^2}{D}, \tag{6}$$

and for this, the rate of loss by diffusion is equal to q ions per c.c. per sec., i.e.

$$\left(\frac{dN}{dt}\right)_{\text{diffusion}} = \tfrac{3}{2}N\frac{D}{l^2}$$

$$= N^2\left(\frac{3D}{2Nl^2}\right), \tag{7}$$

or the contribution made by diffusion to the effective value of α is

$$\frac{3}{2}\frac{D}{Nl^2},$$

and this is the correction which we must subtract from the measured values. It may be noted that it is small for large concentrations, or when the plates are wide apart. For the conditions in these experiments $l = 1{\cdot}5$ cm., and assuming that $D = 0{\cdot}047$ at room temperature and 760 mm. pressure the correction becomes

$$\delta\alpha = \frac{1}{N}3{\cdot}13 \times 10^{-2}$$

and may be neglected for $N = 10^6$.

At 100 mm. pressure $\delta\alpha$ becomes

$$0{\cdot}25 \times 10^{-6},$$

and this correction must be subtracted from the measured value.

5. THE VALUE OF α IN AIR AT ATMOSPHERIC PRESSURE

The recombination coefficient α is evaluated by solution of the well-known differential equation

$$\frac{dN}{dt} = -\alpha N^2.$$

If we write this equation as $\alpha = \dfrac{d}{dt}\left(\dfrac{1}{N}\right)$,

we see that α is the slope of the graph which represents the variation of $1/N$ with time. A graphical analysis of this type was employed in the present experiments.

Fig. 3A is a curve of $1/N$ versus t in the case of air, and was obtained when the ionization was produced by a very short flash of X-rays, the pressure being 760 mm. Hg. The value of α is obtained by graphical differentiation of this, and the result is shown in fig. 3B which is a graph of the variation of α with time after the cessation of the X-rays. It is to be noted that α starts from a very high value initially, but quickly drops to about half its initial value in a time of approximately 0·1 sec., after which it remains constant. This effect has been previously discovered by Marshall, Luhr, and very recently by Gardner (1938) working with pure oxygen. It is due to the fact that at short intervals after the irradiation by X-rays the ions are still distributed in clusters along β-tracks; and therefore a high value of α is observed until the process of diffusion produces a uniform volume distribution of ionization.

It is very illuminating to compare the results of fig. 3B with the value obtained for α by the Rümelin method. The gas is continuously exposed to the X-rays. Let therefore the rate of formation of ions be q pairs per c.c. per sec. If N_0 is the equilibrium value of ion concentration, it is easy to show that α is given by

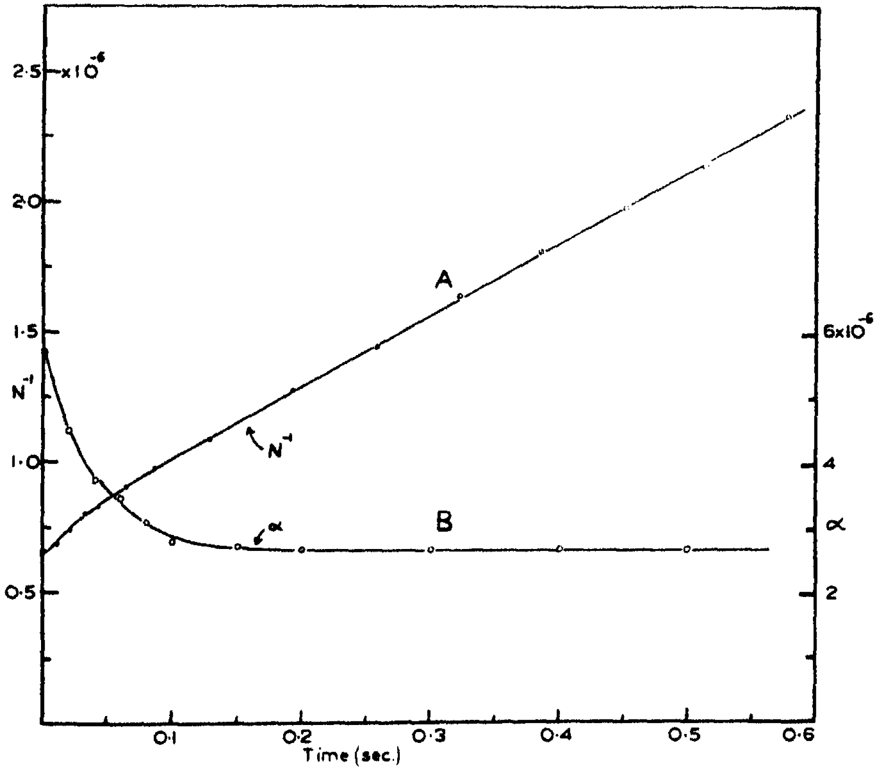


FIG. 3. Curves illustrating the initial high value of α due to columnar and cluster ionization at short time intervals after the cessation of the ionizing rays.

$$\alpha = \frac{q}{N_0^2}; \tag{8}$$

q and N_0 are measured as follows:

With the X-rays on continuously, the time t of application of the collecting field is varied, and a graph plotted of charge collected versus t. If the collecting field is sufficient to draw the saturation current, then the graph is a straight line, and the intercept on the N axis gives the value of N_0, while the slope gives q.

Measurements were made by this method under conditions as nearly as possible identical with those under which the results of fig. 3 were obtained. The value for α arrived at was $5 \cdot 2 \times 10^{-6}$ c.c. per sec. The fact that this is approximately equal to the abnormal high value of α by the flash method at very short time intervals after the cessation of the X-rays is very instructive. It tells us that the results by the Rümelin method are characteristic of conditions where the ion distribution is non-uniform, and that with the gas continuously exposed to X-rays a large proportion of the recombination takes place in the localized regions of high ion concentration such as fresh β-tracks and δ-ray clusters. It is therefore evident that any experiments on the measurement of recombination in the presence of an ionization process are reliable only under special conditions. For example, it would be sufficient if the ionizing agency gave a distribution of ions which was homogeneous throughout the body of the gas, ultra violet light being satisfactory in this respect. In addition, if the processes of diffusion were so effective that the probability of a pair of ions recombining within the cluster where they were formed was small, the Rümelin method could still be applied.

There is also another difficulty which arises in the flash method of determining α for air or oxygen. If, for example, various sets of observations are made in which only the initial intensity of the ionization flash is varied, then on examining the results it is found that:

(i) Each set is consistent in itself, that is after an initial period of 0·1 sec. subsequent to the cessation of the X-rays, α attains a value which thereafter remains constant up to the maximum time studied, or about 0·5 sec.

(ii) There is a remarkable variation of α as the initial intensity of the X-rays is altered.

This phenomenon is clearly illustrated by the results of fig. 4. In this diagram are reproduced some $1/N$ versus t curves for air at 760 mm. pressure and various initial X-ray exposure times. The slopes of these graphs give the values of α, and the manner in which this varies with X-ray exposure is illustrated in Table I.

A similar variation of α with X-ray intensity has been observed by Marshall, Luhr, and Gardner (1938), the latter working with oxygen. Moreover, if instead of varying the X-ray exposure time the hardness of the rays is altered, an even more striking change in α is noted; for example, with a large ionization chamber having a thin cardboard window to admit the X-rays (instead of the pyrex glass wall) the value of α was found to drop to $1 \cdot 65 \times 10^{-6}$ cm.^{3} sec.$^{-1}$. It is evident that everything can be explained by some change in the nature of the gas or the ions, which is produced by the X-rays and will therefore vary with their intensity or exposure time.

TABLE I

N_0 10^6 cm.$^{-3}$	α 10^{-6} cm.^{3} sec.$^{-1}$
1·5	2·65
2·7	2·5
3·1	2·5
4·1	2·3
4·4	2·3

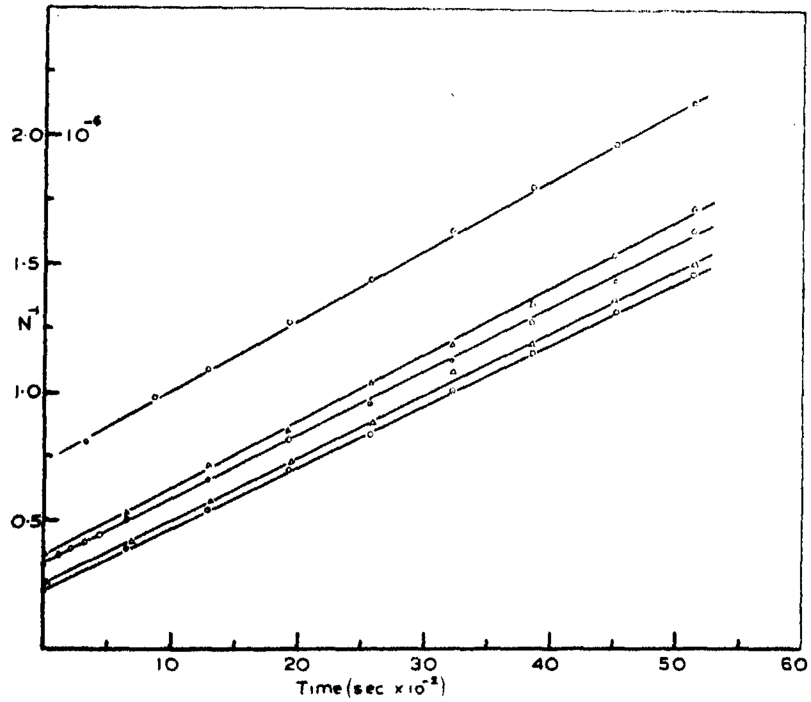


FIG. 4. $1/N$ versus t curves for various values of initial ionization intensity.

Now the most abundant chemical change produced in air by X-rays is the formation of ozone, and this was suspected as leading to the formation of abnormally heavy ions whose number would depend on the amount of ozone present, as also would the value of α recorded. This explanation is amply supported by the following facts: (i) The effect has been found by Gardner to be particularly marked in oxygen. (ii) It is not observed in either nitrogen* or helium, where the X-rays do not form any complex molecules. (iii) Tyndall (1928) has found a similar effect on the ionic mobility, which was reduced, at the short time intervals, by the formation of complex molecules either by X-rays or a corona discharge.

---

* A result of unpublished investigations by G. L. Rogers.

It is most important that this variation of recombination with X-ray exposure time be clearly understood, because on it depends our arrival at a correct value for α in air. The decrease of α as the time of exposure is increased has been ascribed to the fact that a longer exposure time means a larger number of clusters or columns of ionization per unit volume, and therefore a shorter time for the diffusion process to produce uniform distribution of the ionization. If this hypothesis is correct, then the value of α obtained with the highest possible X-ray intensity is obviously the most accurate.

If on the other hand the lowering of α is due to the formation of complex molecules such as ozone or the oxides of nitrogen, one must adopt the results of experiments under conditions where the effect of such a process is least, that is with short exposure times to the X-rays.

The accurately constant value of α observed in these experiments makes it clear that the first hypothesis is incorrect, and that diffusion produces a perfectly homogeneous distribution in about 0·1 sec. We are therefore left with the second explanation to which all the evidence points as the correct one, and accordingly the value of α for the low ion concentration is adopted in all these experiments, with a diffusion correction where necessary.

The result of the adoption of this theory is that the values which will be given for the recombination coefficient in air are considerably greater than those hitherto accepted.

6. THE VARIATION OF α WITH PRESSURE

Attention is now directed to the variation of α with pressure. At pressures of several atmospheres and upwards it is fairly well established that the recombination coefficient is inversely proportional to the gas pressure. This has recently been carefully verified in some admirable work by Mächler (1936), and is in fact in conformity with a theory of Langevin (1903) in which recombination is governed entirely by the ionic mobility constant. At lower pressures, particularly in the region from two atmospheres downwards, the evidence has hitherto been very conflicting, and therefore an investigation of that region was made the basis of many experiments. The experimental procedure employed was the flash method, for the results by this method, as has been explained, are the most instructive and are most easily interpreted.

Fig. 5 gives a few of the $1/N$ versus t curves for air at various pressures. It is to be noticed that these are nearly all quite linear, indicating that the recombination law of decay is followed and that losses by diffusion are small. The effect of loss of ions by diffusion may, however, be noticed at the lower ion concentration ends of some of the curves for the lowest pressures. This is to be expected for two reasons; not only does the rate of diffusion increase as the pressure is reduced, but the necessarily lower value of N increases the ratio of diffusion loss to recombination. The pressure range investigated extends from 100 to 1500 mm. Hg. At pressures below 100 mm. the necessary diffusion correction becomes rather large, and since the expression deduced in § 4 above gives only the approximate correction at best, it is felt that, with the X-ray intensity available, measurements below 100 mm. would not be reliable.

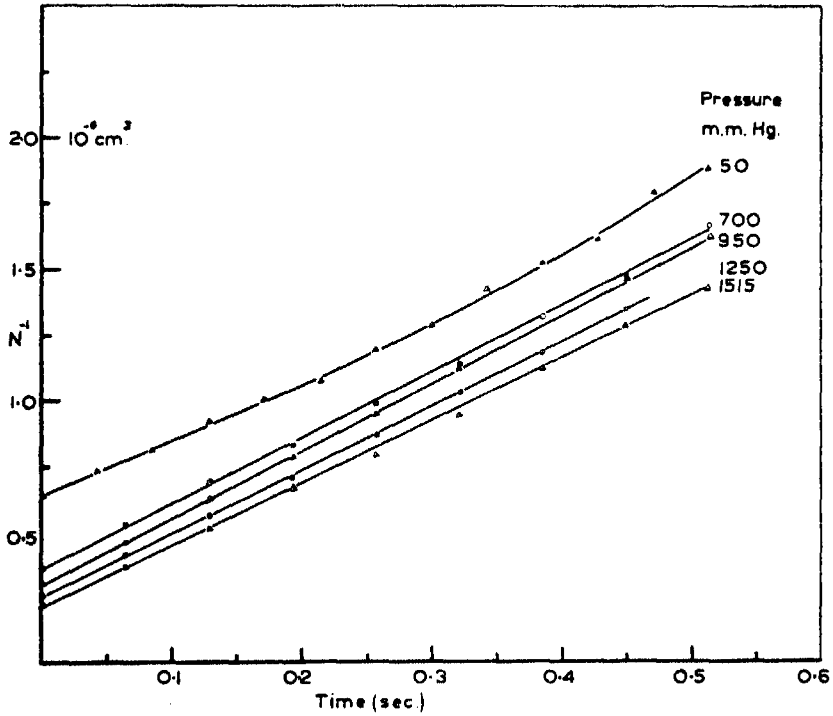


FIG. 5. Sample $1/N$ versus t curves at various pressures.

The results of the measurements of α for the above range in air are given in fig. 6. The experimental points have been corrected for diffusion. The line on the graph gives the variation of α with pressure that is expected on the theory developed by J. J. Thomson (1924). This theory is well known, yielding an expression for α of the form

$$\alpha = \frac{C}{M^{\frac{1}{2}}}(KT)^{-\frac{3}{2}}\,\omega, \tag{9}$$

where C is a constant,

M the average mass of ions,

K the Boltzmann constant, and

ω the chance that one or other of the ions should make a collision with a neutral molecule when the separation of the two is less than or equal to a certain specified distance.

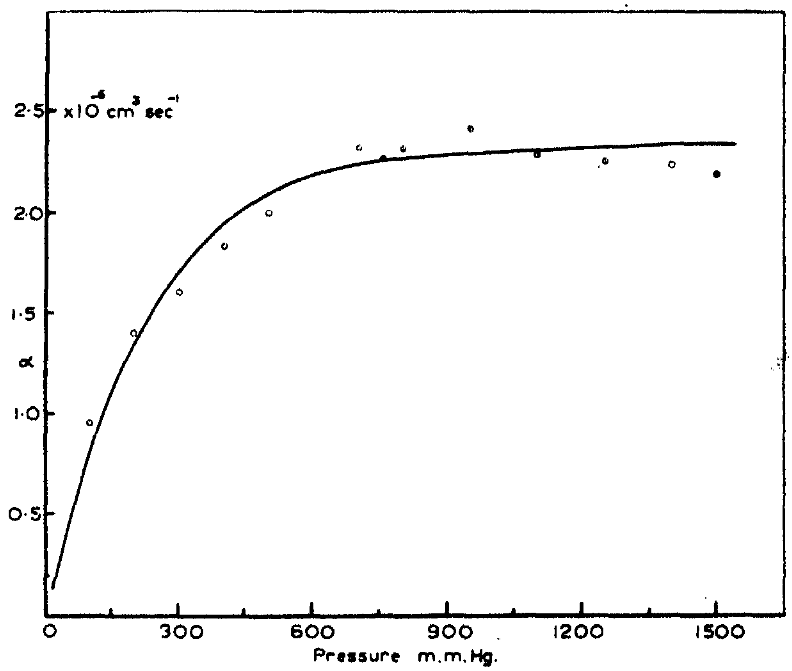


Fig. 6. The dependence of α on pressure. (The points represent experimental determinations; the continuous curve represents the predicted values according to the Thomson theory.)

It is now accepted that this theory represents the essential mechanism of ionic recombination at pressures below 1 atm. For higher pressures, however, it predicts that α should reach a constant value, because ω then becomes equal to unity. This is not substantiated by experimental results, as α ultimately decreases with increasing pressure, having a maximum in the neighbourhood of 1000 mm. Hg. It is therefore evident that at pressures above one or two atmospheres some physical process not accounted for in the Thomson theory begins to assume significance and decreases the value of α.

Fig. 7 is a graph showing the variation of α with pressure over an extended pressure range in air. The present results have been used at pressures between 100 and 1500 mm.; while for high pressures the values of Mächler are given.

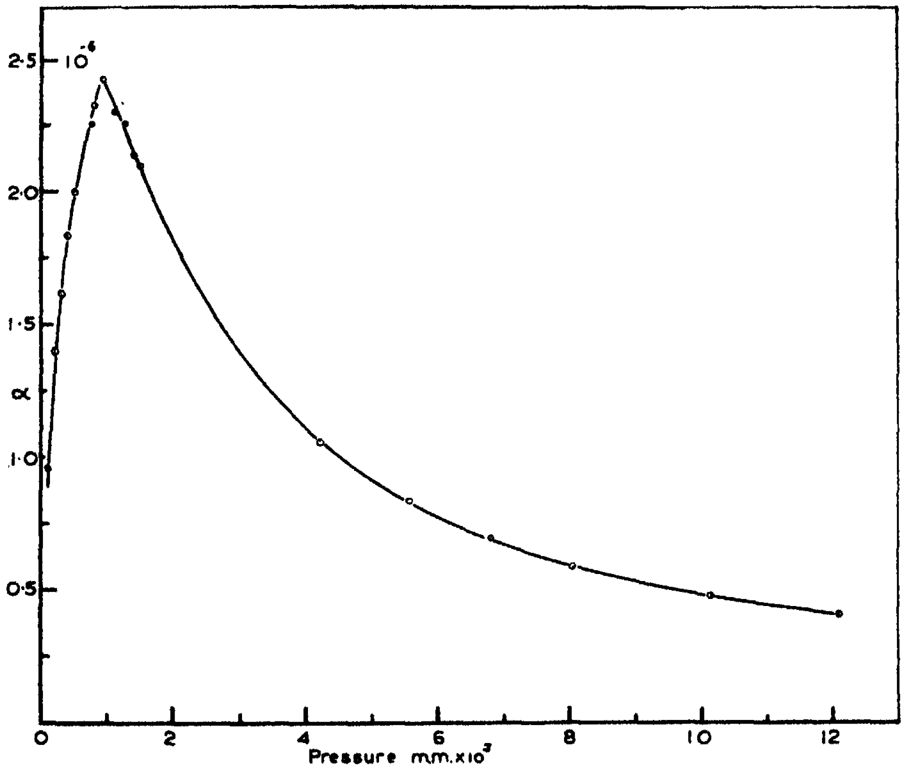


FIG. 7. Graph of α versus pressure for an extended range in air. (The results of the present investigation have been used up to 2 atm. pressure; while for pressures of 5 atm. upwards the determinations are due to Mächler.)

We may summarize the pressure variation of α in air as follows:

(i) At pressures from a few mm. up to 760 mm. Hg. α varies in a manner which is accurately described by the Thomson theory; that is to say, α is proportional to pressure at low values of p, and approaches a limiting value in the region of 1000 mm. Hg.

(ii) Above this pressure, and up to 10 atm., there is a transition period in which α decreases with increasing pressure in a manner which is not satisfactorily explained by any existing theory.

(iii) For pressures above 10 atm., α is approximately proportional to the reciprocal of pressure in accordance with the Langevin mobility formula.

7. RECOMBINATION AND THE IONIZATION BALANCE IN THE LOWER ATMOSPHERE

When a gas is subject to continuous ionization, an equilibrium is soon established between the rate of formation of ions and the rate of loss by recombination, diffusion, and any other process such as the removal of ions by applied electric fields. When recombination alone is operative as the destructive process, the solution of the ion balance equation is exceedingly simple, and the equilibrium concentration, N_0 ions per c.c., can be easily expressed in terms of the rate of ion formation q per c.c. per sec. and the recombination coefficient α. Experimental determinations of N_0 have been made by Gish and Sherman.* They constructed an automatic recording apparatus which was taken aloft on the stratosphere flight of *Explorer II* at South Dakota on 11 November 1935. With this, continuous measurements were made of the electrical conductivity of the atmosphere both on the ascent and descent. For comparison with the direct measurements, they have calculated the value of N_0 and the conductivity, using the old values of α, and the measurements of q made by Regener with sounding balloons at approximately the same latitude as Dakota. Very poor agreement was found between these calculated values of conductivity and the observations of the stratosphere flight. However, since it is now believed that the values assumed for α and its variation with pressure and temperature were inexact, a similar calculation is now made using the new information concerning α as given above.

We assume that cosmic rays are the only source of ionization at heights between 3 and 17 km. above sea-level; and we accept the measurements made by Regener of the number q of pairs of ions produced per c.c. per sec. We also assume that the mobility μ varies directly as the absolute temperature T and inversely as the pressure p.

Thus
$$\mu = \mu_0 \frac{T}{273} \frac{760}{p}, \tag{10}$$
where μ_0 is the mobility at N.T.P.

---

[*] See Stevens and Anderson (1938).

We consider only the case of positive ions, as the positive conductivity measurements of the stratosphere flight are the more complete.

We have

$$\text{Conductivity} = e\mu N_0$$

$$= e\mu_0 \frac{T}{273} \frac{760}{p} N_0. \tag{11}$$

But

$$N_0 = \sqrt{\frac{q}{\alpha}}$$

$$= \left\{ \frac{q_0}{\alpha_{273}} \frac{p}{760} \left(\frac{T}{273} \right)^{\frac{3}{2}} \right\}^{\frac{1}{2}}, \tag{12}$$

where q_0 is the value of q reduced to N.T.P. and α_{273} is the recombination coefficient for the particular value of p but corrected to 0° C. The complete expression for conductivity is, therefore,

$$K = \frac{e\mu_0 T^{\frac{7}{4}}}{p^{\frac{1}{2}}} \frac{(760)^{\frac{1}{2}}}{(273)^{\frac{7}{4}}} \sqrt{\frac{q_0}{\alpha_{273}}}. \tag{13}$$

From this expression, and substituting the values of q_0 given by Regener, and the values of α measured in these experiments, the conductivity is determined as a function of pressure, and therefore of the height above sea-level.

The curve of fig. 8 is a graphical plot of conductivity versus height obtained as above, while the experimental points are some representative values from Gish and Sherman's automatic records of the stratosphere flight.

In the region between 15 and 17 km., where the agreement is not quite accurate, the observed values were not directly recorded, but derived from the negative conductivity; so there is the possibility of a slight error in this region.

However, the general accuracy of the agreement indicates how precise is our knowledge of the magnitudes of these fundamental electrical processes occurring in the lower atmosphere.

It is of interest to note that there is no indication here of the existence, at any height below 20 km., of regions of sufficient ion density to reflect radio waves to any measurable degree.

In conclusion I wish to express my gratitude to Professor E. V. Appleton, F.R.S. for his continual interest and many suggestions which have contributed materially to the completion of the experiments here described.

The work was undertaken as part of the programme of the Radio Research Board of the Department of Scientific and Industrial Research.

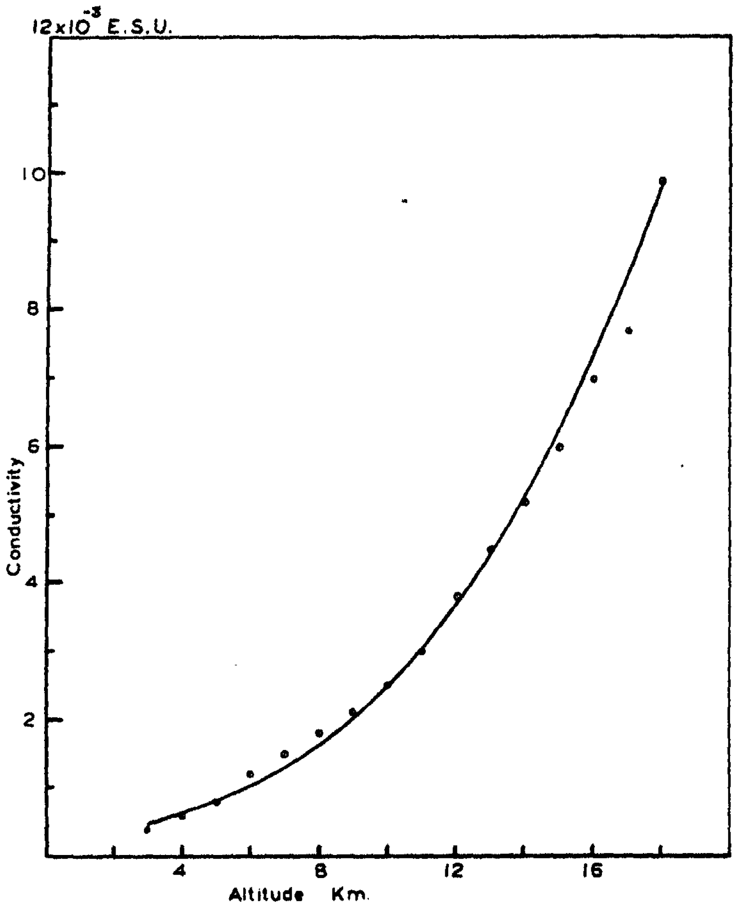


FIG. 8. Graph of electrical conductivity versus height in the lower atmosphere. (The points represent the experimental values of Gish and Sherman; the continuous curve represents the values calculated, using the new values of α, and known values of cosmic radiation intensity.)

SUMMARY

An experimental study has been made of the recombination of ions in air. It is found that, up to pressures of 1 atm., the recombination coefficient varies with pressure in a manner which is accurately predicted by the theory of J. J. Thomson; that is, the recombination coefficient varies proportionally with the pressure up to about 100 mm., thereafter varying less rapidly as the pressure is increased, until, at 1000 mm., the recombination coefficient becomes pressure-independent.

The apparent variation of the recombination coefficient with the ionization intensity, noted by other workers, has been confirmed and explained as being due to the formation, by the ionizing rays, of complex molecules such as ozone, new experimental evidence in favour of this explanation being adduced.

A number of subsidiary experiments have yielded detailed information concerning the nature of the recombination process, from which the inaccuracies in previous work on this subject may be appreciated.

Using the new values of the recombination coefficient for air, it has been possible to calculate the equilibrium ion density and the electrical conductivity at various levels in the lower atmosphere. For this purpose the rates of ion production measured by Regener by means of sounding balloons have been employed. The results so obtained are shown to agree very closely with the direct measurements of the atmospheric conductivity made in the stratosphere ascent of *Explorer II* at South Dakota by Stevens and Anderson in 1935.

REFERENCES

Appleton 1928 *J. Instn. Elect. Engrs*, **66**, 875.
Gardner 1938 *Phys. Rev.* **53**, 75.
Gish and Sherman 1936 *Bull. Int. Ass. Terr. Mag. Elect.* No. 10, p. 382.
Langevin 1903 *Ann. Chim. Phys.* **28**, 433.
Lenz 1932 *Z. Phys.* **76**, 660.
McClelland 1898 *Phil. Mag.* **46**, 29.
Mächler 1936 *Phys. Z.* **37**, 211.
Rümelin 1914 *Ann. Phys., Lpz.*, **43**, 821.
Rutherford 1897 *Phil. Mag.* **44**, 422.
Schemel 1928 *Ann. Phys., Lpz.*, **85**, 137.
Stevens and Anderson 1938 *Nature, Lond.*, **141**, 270.
Thirkill 1913 *Proc. Roy. Soc.* A, **88**, 488.
Thomson, J. J. 1924 *Phil. Mag.* **47**, 337.
— 1928 "Conduction of Electricity through Gases", **1**, 3rd Ed. Camb. Univ. Press.
Townsend 1899 *Philos. Trans.* A, **193**, 144.
— 1900 *Philos. Trans.* A, **193**, 157.
Tyndall and others 1928 *Proc. Roy. Soc.* A, **121**, 172, 185.

On the Hamilton-Jacobi theory and quantization of a dynamical continuum

By P. Weiss, Ph.D., *Downing College, Cambridge*

(*Communicated by R. H. Fowler, F.R.S.—Received* 1 *September* 1938)

1. Introduction

The quantization of the dynamics of point systems is closely connected with the Hamilton-Jacobi theory of the calculus of variations for *simple* integrals, the latter being a suitable mathematical formalism for describing the classical laws of point dynamics. This connexion is evident from the fact that, in order to quantize a dynamical system, one has first to know what the *pairs of canonically conjugate variables* are, and also from the fact that the quantum laws, if expressed in terms of commutation brackets, have exactly the same form as the classical laws, if expressed in terms of *Poisson brackets*.

In dealing with the quantization of the dynamics of continuous media, e.g. the electromagnetic and other fields, which has been developed along different lines, following Heisenberg and Pauli (1929), it seems tempting to try to base the method of quantization on an extended Hamilton-Jacobi theory of the calculus of variations for *multiple* integrals, the latter being the appropriate formalism for describing most relevant systems of continuum physics.

In a previous paper (Weiss 1936), referred to as I, such an attempt has been made, by extending the notion of pairs of conjugate variables. That attempt led to quantum relations on an arbitrary space-like section in space-time, provided that one *postulated* that, if the space-like section becomes especially a space section, the quantum relations should go over into those of Heisenberg and Pauli.

In the present paper the notion of Poisson brackets will be extended as well, and it will be *deduced* that the classical laws of a dynamical continuum (which satisfies a variation principle), expressed in terms of Poisson brackets, have exactly the same form as the quantum laws arrived at in I. The procedure of quantizing a dynamical continuum therewith becomes exactly the same as the procedure of quantizing point dynamics, and it will yield all physically relevant results of I.

The main part of this paper will be concerned with classical considerations, i.e. the Hamilton-Jacobi theory of continua leading from Euler's partial differential equations to Poisson brackets. The transition to the quantum theory of continua will then be a matter of a few lines. The argument, which is modelled on the Hamilton-Jacobi theory of mechanics, starts from the "complete variation" and "boundary formula", like the argument of I, and will then lead on to an extension of Poincaré's "relative integral invariant", from there to Lagrange brackets and then to Poisson brackets.

To obtain Poincaré's integral invariant and the subsequent notions, it will be necessary to use a representation of all quantities in terms of *initial data*. The use of such a representation presupposes that the state of the dynamical continuum in space-time is uniquely determined by initial data (plus boundary data in space) and that it depends continuously on these data. In other words, the initial value problem must be *correctly set*.* This is true only for theories describing a *dynamical* continuum, not, for instance, for equilibrium theories. Thus we have the satisfactory feature that the method of quantization is for purely mathematical reasons restricted to dynamical theories, as it should be on obvious physical grounds.

The initial data need not be given on a space section ($t=\text{const.}$), but may be given on any space-like section. Mathematically, this is well known from the theory of partial differential equations. The physical implications of this fact are that the present formalism will be relativistically invariant throughout. This is an advantage as compared with the work of Heisenberg and Pauli, where a space section was fixed once and for all, in consequence of which the relativistic invariance of the result was not apparent and had to be proved separately.

Another implication relates to the notion of "state" in continuum dynamics. The word "state" in the classical sense (cf. Cartan 1922, p. 4) is synonymous with "correctly set data", and it is clear from the above remarks that it will be relativistic. It is also clear, from the meaning of the word "data", that though referring only to a space-like cross-section, it describes the motion of the dynamical continuum throughout space-time. The same is true for the word "state" in the quantum sense (Dirac 1935, ch. I), meaning the maximum information of non-contradictory data relating to the dynamical system. For the fundamental difference between the classical and the quantum sense of the word state does not affect the question of its relativistic invariance. Thus by allowing space-like sections as bearers of the data, and not restricting oneself to space sections of constant time, the two meanings of the word state, distinguished by Dirac both in classical and quantum physics, can be reconciled with each other.

---

* For a detailed discussion of the notion of "correctly set data" in the theory of partial differential equations, cf. Hadamard (1923, 1932, ch. I) and Courant-Hilbert (1937, chs. III, V).

Some basic considerations of I will be repeated. As in I, the continuum will be n-dimensional and will be described by ν dependent variables. The notions of "space-like" and "time-like" can still be defined mathematically for an n-dimensional space-time, through the theory of differential equations of hyperbolic type. This generality serves to accentuate the mathematical character of our procedure. The comparison with the case of mechanics, $n = 1$, will be discussed at different stages of the argument. The mathematical tool applicable to the general case, $n > 1$, is furnished by the theory of functionals.*

The present considerations are not immediately applicable to physics, since all physical continuum theories, like electrodynamics, have special features requiring special discussions.† We shall deal with these questions in a later communication where the present considerations will serve as a suitable basis.

2. ACTION INTEGRAL

We consider a system with n independent variables x^i and ν dependent variables z^α, whose classical laws satisfy a variation principle. The derivatives $\frac{\partial z^\alpha}{\partial x^k}$ will be denoted by z^α_k. Let

$$\mathsf{I} = \int_D^{(n)} \mathscr{L}\{x^i, z^\alpha(x), z^\alpha_k(x)\}\, dx \quad \begin{pmatrix} i, k = 1, 2, \ldots, n \\ \alpha = 1, 2, \ldots, \nu \end{pmatrix} \tag{2·1}$$

be the action integral.‡ It is an n-fold integral extended over a domain D of the n-dimensional "x-space". dx stands for the product $dx^1 dx^2 \ldots dx^n$. The boundary of D, which forms an $(n-1)$-dimensional closed manifold, will be denoted by S and be referred to $n-1$ parameters u^r. The integrand $\mathscr{L}$ is the Lagrangian function which characterizes the particular system under consideration.

---

* Cf. Volterra (1913, 1930, 1937), De Donder (1935), Prange (1915), and Juvet (1926). The present considerations resemble more closely Prange's work than that of any other author. Many of the notions used here have been defined by Prange for the case of a two-dimensional continuum. He did not, however, consider the question of correctly set data, so that his arguments referring to Poincaré's integral invariant, etc. are merely formal and his analogies with mechanics not quite correct.

Poisson brackets have first been introduced in the theory of functionals by Volterra (1913, p. 74), but without reference to the calculus of variations.

† In I, § 5, the special features of electrodynamics and their consequences have been discussed in detail.

‡ Throughout this paper Greek indices will vary from 1 to ν, Latin indices $i, j, k, \ldots$ from 1 to n, Latin indices $r, s, \ldots$ from 1 to $n-1$. We shall use the customary summation convention.

In the case $n=1$ and if x denotes the time, we have a mechanical system of ν degrees of freedom. The boundary S degenerates in this case into two distinct points and the action integral becomes

$$\mathsf{I} = \int_1^2 \mathscr{L}\{x, z^\alpha(x), z'^\alpha(x)\}\,dx. \tag{2·1a}$$

This difference between the cases $n=1$ and $n>1$ is responsible for some difficulties which arise if one tries to generalize the theory of Hamiltonian mechanics to the case of continuous media.

It will sometimes be convenient to picture the functions $z^\alpha(x)$ as an n-dimensional surface in the $(n+\nu)$-dimensional "(x,z)-space". Its $(n-1)$-dimensional closed boundary may still be referred to the parameters u^r of S; for S is the projection of that boundary into the x-space.

3. Complete variation

We define variations η^α by varying the functions $z^\alpha(x)$,

$$\bar{z}^\alpha(x) = z^\alpha(x) + \epsilon\,\eta^\alpha(x), \tag{3·1}$$

and variations δx^i, δz^α by

$$\bar{x}^i = x^i + \epsilon\,\delta x^i(x), \quad \bar{z}^\alpha(\bar{x}) = z^\alpha(x) + \epsilon\,\delta z^\alpha(x), \tag{3·2}$$

ϵ being an infinitesimal parameter.

These variations are connected by the relation

$$\delta z^\alpha = \eta^\alpha + z^\alpha_j\,\delta x^j. \tag{3·3}$$

The derivatives $\dfrac{\partial \eta^\alpha}{\partial x^k}$ will be denoted by η^α_k.

If we picture our quantities in the (x,z)-space, the variations δx^i, δz^α are more fundamental than the η^α, because $(\epsilon\,\delta x^i, \epsilon\,\delta z^\alpha)$ is the displacement of the point (x^i, z^α) in the (x,z)-space.

By the "*complete variation*" $\delta\mathsf{I}$ we understand the variation of the action integral I due to the variations (3·1) and (3·2). We obtain for it

$$\delta\mathsf{I} = \int_D^{(n)} \left(\frac{\partial \mathscr{L}}{\partial z^\alpha}\,\eta^\alpha + \frac{\partial \mathscr{L}}{\partial z^\alpha_k}\,\eta^\alpha_k\right) dx + \oint_S^{(n-1)} \mathscr{L} N_i\,\delta x^i\,du,$$

where the N_i are the $(n-1)$-rowed minors of the Jacobian matrix*

$$\left\| \frac{\mathfrak{d}x^i}{\mathfrak{d}u^r} \right\|;$$

they form a vector field on S, normal to S. du stands for the product $du^1\, du^2 \ldots du^{n-1}$.

With the help of the formula

$$\frac{\partial \mathscr{L}}{\partial z^\alpha_k}\eta^\alpha_k = -\left(\frac{\mathfrak{d}}{\mathfrak{d}x^k}\frac{\partial \mathscr{L}}{\partial z^\alpha_k}\right)\eta^\alpha + \frac{\mathfrak{d}}{\mathfrak{d}x^k}\left(\frac{\partial \mathscr{L}}{\partial z^\alpha_k}\eta^\alpha\right)$$

we perform an integration by parts, using Green's theorem, and introduce on the boundary the variations δx^i, δz^α instead of η^α, by means of (3·3). $\delta \mathsf{I}$ then becomes

$$\delta \mathsf{I} = \int_D^{(n)} \left(\frac{\partial \mathscr{L}}{\partial z^\alpha} - \frac{\mathfrak{d}p^l_\alpha}{\mathfrak{d}x^l}\right)\eta^\alpha dx + \oint_S^{(n-1)} (X_i \delta x^i + P_\alpha \delta z^\alpha)\, du, \tag{3·4}$$

where we have introduced the notations†

$$p^i_\alpha = \frac{\partial \mathscr{L}}{\partial z^\alpha_i}, \quad U^i_k = \mathscr{L}\delta^i_k - p^i_\beta z^\beta_k, \tag{3·5}$$

$$P_\alpha = p^l_\alpha N_l, \quad X_k = U^l_k N_l. \tag{3·6}$$

The ν coefficients of η^α under the n-fold integral on the right-hand side of (3·4) are the Eulerian expressions. If they vanish, we have the Eulerian equations

$$\frac{\partial \mathscr{L}}{\partial z^\alpha} - \frac{\mathfrak{d}p^l_\alpha}{\mathfrak{d}x^l} = 0. \tag{3·7}$$

In physics these equations constitute the classical laws. Their solutions are the "extremals" of the calculus of variations.

In the case of mechanics, $n = 1$, the Eulerian equations are the equations of motion. In the case of a continuum, $n > 1$, they are one set of field equations, another set being given by the integrability conditions which the $z^\alpha_k(x)$ must satisfy in order to be representable as derivatives $\frac{\mathfrak{d}z^\alpha}{\mathfrak{d}x^k}$; these read

$$\frac{\mathfrak{d}z^\alpha_k}{\mathfrak{d}x^i} - \frac{\mathfrak{d}z^\alpha_i}{\mathfrak{d}x^k} = 0. \tag{3·8}$$

From now onwards we shall assume that (3·7) and (3·8) are satisfied.

---

* The quantities N_i have in I been denoted by X_i. We have changed our notation in a few details in the hope that the present notation will be more suggestive.

† The quantities X_i have in I been denoted by σ_i and the P_α by π_α.

4. BOUNDARY FORMULA AND CONJUGATE VARIABLES

It follows from (3·4) that extremals, instead of being characterized by (3·7), may alternatively be characterized by the equation

$$\delta \mathsf{I} = \oint_S^{(n-1)} (X_i \delta x^i + P_\alpha \delta z^\alpha)\, du. \tag{4·1}$$

We shall call this equation the "*boundary formula*".[*] The action integral I may be considered as a functional of the boundary—$x^i(u)$, $z^\alpha(u)$—in the (x, z)-space and the boundary formula (4·1) as its differential. X_i and P_α are therefore the functional derivatives of I with respect to x^i and z^α respectively, taken at a point (u) of the boundary. By this property X_i and P_α are defined as the conjugates to x^i and z^α so that we have the "*pairs of conjugate variables*"

$$(X_i, x^i), \quad (P_\alpha, z^\alpha). \tag{4·2}$$

In the case of mechanics the boundary formula becomes

$$\delta \mathsf{I} = \left| \mathscr{H} \delta x + p_\alpha \delta z^\alpha \right|_1^2, \tag{4·1a}$$

where we have put[†]

$$p_\alpha = \frac{\partial \mathscr{L}}{\partial z'^\alpha}, \quad \mathscr{H} = \mathscr{L} - p_\beta z'^\beta \tag{3·5a}$$

and inverted the sign of the variations at the point 1. $\mathscr{H}$ represents the negative energy of the system. It will be noticed that (3·5) and (3·6) *both* reduce to (3·5*a*).

I degenerates into a "two-point function" of the co-ordinates (x^i, z^α) of the two boundary points in the (x, z)-space. The pairs

$$(\mathscr{H}, x), \quad (p_\alpha, z^\alpha) \tag{4·2a}$$

are the well-known conjugate variables of Hamiltonian mechanics. This fact shows that our present considerations may be considered as a generalization of Hamiltonian mechanics.

It should be noted that in the general case, $n > 1$, the conjugates of x^i and z^α respectively depend not only on a point in the x-space, like the z^α, but also on the *direction* of a surface element passing through this point. This fact will be important in our subsequent considerations.

---

[*] This term is the translation of the French "formule aux limites". Cf. Hadamard (1910, pp. 142–51).

[†] Our definition of $\mathscr{H}$ differs from the usual one in sign.

5. NATURAL CO-ORDINATE SYSTEM AND LEGENDRE TRANSFORMATION

A "*natural co-ordinate system*" is defined in the x-space and is adapted to the boundary S in such a way that on S the first co-ordinate, say w, is normal to S and the other $n-1$ co-ordinates are the parameters u^r of S. We shall use the notation

$$z'^{\alpha} = \frac{\mathfrak{d} z^{\alpha}}{\mathfrak{d} w}, \quad \tilde{z}^{\alpha}_{r} = \frac{\mathfrak{d} z^{\alpha}}{\mathfrak{d} u^{r}}. \tag{5·1}$$

For the conjugates P_α, X_k, defined by (3·6), we obtain in a natural co-ordinate system

$$P_\alpha = \frac{\partial \mathscr{L}}{\partial z'^{\alpha}}; \quad X_1 = \mathscr{L} - P_\beta z'^{\beta}, \quad X_{1+r} = -P_\beta \tilde{z}^{\beta}_{r}. \tag{5·2}$$

Let us put $$\mathscr{H} = \mathscr{L} - P_\beta z'^{\beta}, \quad h_r = -P_\beta \tilde{z}^{\beta}_{r}. \tag{5·3}$$

The boundary formula (4·1) may now be written as

$$\delta I = \oint_{S}^{(n-1)} (\mathscr{H} \delta w + h_r \delta u^r + P_\alpha \delta z^\alpha)\, du. \tag{5·4}$$

By means of the *Legendre transformation* (Courant-Hilbert 1937, p. 26)

$$z'^{\alpha} \rightarrow P_\alpha, \quad \mathscr{L}(z^\alpha, \tilde{z}^{\alpha}_{r}, z'^{\alpha}) \rightarrow \mathscr{H}(z^\alpha, \tilde{z}^{\alpha}_{r}, P_\alpha) \tag{5·5}$$

we replace the derivatives z'^{α} normal to S by the P_α, and the classical laws can now be written in the "*canonical form*"

$$\frac{\partial \mathscr{H}}{\partial z^\alpha} - \frac{\mathfrak{d}}{\mathfrak{d} u^r} \frac{\partial \mathscr{H}}{\partial \tilde{z}^{\alpha}_{r}} - \frac{\mathfrak{d} P_\alpha}{\mathfrak{d} w} = 0, \quad \frac{\partial \mathscr{H}}{\partial P_\alpha} + \frac{\mathfrak{d} z^\alpha}{\mathfrak{d} w} = 0. \tag{5·6}$$

In the case $n=1$ the question of the natural co-ordinate system does not arise, $\mathscr{H}$ goes over into the Hamiltonian of mechanics, and the canonical equations take the well-known form

$$\frac{\partial \mathscr{H}}{\partial z^\alpha} - \frac{dp_\alpha}{dx} = 0, \quad \frac{\partial \mathscr{H}}{\partial p_\alpha} + \frac{dz^\alpha}{dx} = 0. \tag{5·6a}$$

Thus the function $\mathscr{H}$ defined by (5·3) appears as the most natural generalization of the Hamiltonian of mechanics, though other generalizations have also been given.*

Although $\mathscr{H}$ has been introduced as a component—X_1—of a vector in the natural co-ordinate system, it has an invariant significance, because for a given surface element of S the co-ordinate w normal to S has an invariant significance, and $\mathscr{H}$ is the conjugate to this co-ordinate w.

---

* The best known one is $\mathscr{H} = \mathscr{L} - p^{1}_{\beta} z^{\beta}_{1}$, first given by Volterra.

6. BOUNDARY AND INITIAL DATA. THE HYPERBOLIC TYPE

So far we have considered the extremal surfaces $z^\alpha(x)$ as given by their *boundary* $x^i(u)$, $z^\alpha(u)$ in the (x, z)-space, or in other words, by prescribing the "*boundary data*" $z^\alpha(u)$ on the whole boundary S in the x-space. Generally speaking, one understands by boundary data the prescribing of *one* quantity, e.g. z^α *or* P_α, on the whole of S.

For many considerations, however, in particular for deriving Poincaré's integral invariant, Lagrange brackets and Poisson brackets, it is necessary to characterize extremals by "*initial data*". Here we understand by initial data that on a part of the boundary S *two* data, e.g. z^α *and* P_α, are prescribed, while on another part of S no datum is prescribed.

In mechanics an extremal may in general be given by initial data equally well as by boundary data. In either case it is uniquely determined by its data* and depends continuously on them; the same is therefore true of the action integral I. There is thus no difficulty in passing over from the one standpoint to the other, as is done in the Hamilton-Jacobi theory of mechanics.

In a continuum theory the position is more complicated. One can distinguish between different types of systems according to what kind of data are "*correctly set*". This term has been explained in the introduction. There is one type of systems, the so-called "*elliptic type*", for which boundary data are in general correctly set, but initial data are not. In continuum physics all *equilibrium problems* belong to this type.

All physical problems of continuum *dynamics* belong to another type, called the "*hyperbolic type*". Before we can state precisely what kind of data are correctly set for this type, we have to give some preliminary explanations. For a detailed treatment we refer to the works of Hadamard and Courant, quoted in the introduction. There exists for the hyperbolic type a *real* $(n-1)$-dimensional "*characteristic cone*" (in physics called "light cone") at each point of the x-space, which gives rise to a classification of a given closed boundary S into different regions. For $n > 2$ there are regions on S, called *space-like,* where the characteristic cone at a given point P of S intersects S only in P itself and in no other point of a suitable neighbourhood of P; there are other regions on S, called *time-like,* where the cone at P intersects S in an infinity of points in any given neighbourhood of P. The frontier between space-like and time-like regions on S is an $(n-2)$-dimensional manifold of points whose characteristic cones *touch* S. For $n = 2$,

---

* The case where this is not true, e.g. "conjugate points", arises from an exceptional choice of data.

there are still the same different regions on S, but their roles of being space- or time-like may be interchanged, the characteristic cone degenerating into a pair of lines passing through P.

In physics, the history of a domain of space during an interval of time is described by a region D in space-time ($n=4$) of cylindrical shape. The boundary S of D consists of two bases, which are space-like, and of one curved part representing the history of the boundary of the domain of space, which is time-like. The principle of general relativity permits all those transformations of space-time which preserve the character of a region of being space- or time-like. Any closed surface S consisting of two distinct space-like regions, S_1 and S_2 say, and of a time-like region, joining S_1 with S_2, is relativistically equivalent to a cylinder of the kind described above; it is therefore admissible as a boundary for a problem of continuum dynamics.

For all problems of the hyperbolic type boundary data throughout S are in general not correctly set. The only correctly set data are: *initial data, z^α and P_α, on one of the two space-like regions of S, S_1 say; boundary data, z^α or P_α, on the time-like region; no data on the remaining space-like region, S_2 say.* Such data are called "*mixed data*".

For $n > 3$, there exist still other types of systems, called "ultrahyperbolic", for which neither boundary data nor mixed data are in general correctly set. No physical problem belongs to such a type.

From now on we shall deal with the hyperbolic type only.

7. INTERPRETATION OF THE BOUNDARY FORMULA FOR THE HYPERBOLIC TYPE

It is clear from the preceding considerations that the notions of integral invariant, Lagrange brackets and Poisson brackets can be extended only to continuous systems of the hyperbolic type and only as far as the initial data, i.e. the data on a space-like region are concerned. This, however, is precisely what physics requires.

On the other hand, the boundary formula (4·1) has only a formal meaning in the case of continuum dynamics, since the functional I does not depend uniquely and continuously on the boundary data. The conclusion which we have drawn from it, i.e. which quantities are pairs of conjugate variables, is also of a merely formal kind and is therefore valid for any type of system. But for our further considerations we have to adapt the boundary formula to the fact that I is now a proper functional of the mixed data and not of the boundary data on S.

It would be difficult to obtain an expression for the general functional differential of I, describing the variation of I corresponding to the variations of the mixed data. For our purpose, however, we need only study the special case where the boundary data on the time-like region are kept fixed and the initial data on the space-like region S_1 alone are varied. The boundary formula may then be written as

$$\delta \mathsf{I} = \left| \int^{(n-1)} (X_i \delta x^i + P_\alpha \delta z^\alpha)\, du \right|_{S_1}^{S_2}, \tag{7·1}$$

where we have inverted the sign of the variations on S_1. This formula should be compared with the corresponding formula (4·1 *a*) for the case of mechanics.

In (7·1), the variations on S_2 are determined by the initial data on S_1 and their variations. The right-hand side of (7·1) is, therefore, not a proper differential of the functional I considered as a functional of the initial data. In the following section we shall turn this fact to our advantage.

The fixed boundary data on the time-like region represent spatial boundary conditions holding throughout time. Their physical significance depends on the particular problem under consideration. For vibrations of a string or of a membrane they are conditions imposed on the boundary of the string or membrane (as is obvious). For a pure radiation field of finite extension, enclosed between walls, they represent the optical properties of the walls. For a radiation field extending through the whole of space, they represent conditions for the behaviour of the field at infinity in space. If the field contains charged particles, forming singular time-like world-lines extending in time (more precisely: proper-time) from the infinite past to the infinite future, one may exclude these singular lines by small tubes. One then obtains a multiply connected space-time, for which the boundary data on the time-like region consist not only of conditions at infinity but also of conditions on the surfaces of the small tubes. These latter conditions describe the behaviour of the singularities.

We need not deal with these conditions in the present work, since they are not of the nature of initial conditions and thus do not affect the Hamilton-Jacobi theory and quantization of a dynamical continuum, considered by itself. They may be of importance if one considers interaction processes, for instance the interaction of a radiation field with charged particles.

8. Integral invariant and Lagrange brackets

Let us consider a one-parametric *closed* family of extremals, labelled by a parameter λ, $0 \leqq \lambda \leqq 1$, such that the time-like parts of the boundaries are the same for the whole family. The initial data on S_1 then become functions of λ,

$$S_1 : x^i = x^i(u,\lambda), \quad z^\alpha = z^\alpha(u,\lambda), \quad P_\alpha = P_\alpha(u,\lambda). \tag{8·1}$$

The fact that the family is closed is expressed by the relations

$$x^i(u,0) = x^i(u,1), \quad z^\alpha(u,0) = z^\alpha(u,1), \quad P_\alpha(u,0) = P_\alpha(u,1).$$

We thus obtain a tube of extremals in the (x,z)-space, which is completely determined by the tube of $(n-1)$-dimensional surfaces (8·1) and by the constant time-like boundary part. The functional I thus becomes a function of the parameter λ.

Let us now apply the modified boundary formula (7·1) to this tube by putting

$$\delta x^i = \frac{dx^i}{d\lambda} d\lambda, \quad \delta z^\alpha = \frac{dz^\alpha}{d\lambda} d\lambda$$

and then integrate it round the tube, i.e. from 0 to 1. The result must be zero since

$$\mathsf{I}(0) = \mathsf{I}(1).$$

Hence we obtain

$$\int_0^1 d\lambda \int_{S_1(\lambda)}^{(n-1)} \left(X_i \frac{dx^i}{d\lambda} + P_\alpha \frac{dz^\alpha}{d\lambda} \right) du = \int_0^1 d\lambda \int_{S_2(\lambda)}^{(n-1)} \left(X_i \frac{dx^i}{d\lambda} + P_\alpha \frac{dz^\alpha}{d\lambda} \right) du. \tag{8·2}$$

The formula (8·2) holds for any two tubes of $(n-1)$-dimensional space-like surfaces $S_1(\lambda)$ and $S_2(\lambda)$, provided that they both lie on the same tube of extremals. Hence the expression in (8·2) is invariant under a displacement along the tube. We have thus obtained a generalization of the notion of "*relative integral invariant*".*

$$\int_0^1 d\lambda \int_{S_1(\lambda)}^{(n-1)} \left(X_i \frac{dx^i}{d\lambda} + P_\alpha \frac{dz^\alpha}{d\lambda} \right) du. \tag{8·3}$$

---

* Cf. Cartan (1922, pp. 4, 9), Whittaker (1927, p. 272). Recently Born (1934) has attempted to derive a quantization method from Hilbert's "*independent integral*" (Hilbert 1900, 1905). Hilbert's integral can be given the same form as that of (8·3), but it applies to *selections* (Hilbert calls them "fields") of extremals such that through each point in the (x, z)-space there passes one and only one extremal. Hilbert's "*independence theorem*" is an immediate consequence of the invariance of (8·3). For if a field of extremals is especially chosen in such a way that the closed integral (8·3) vanishes for all tubes $S_1(\lambda)$ on a cross-section of the (x, z)-space, it must, owing to its invariance, do so throughout the field. This is Hilbert's theorem.

Since a displacement within the same extremal gives only a trivial contribution, we may without loss of generality assume that $\frac{dx^i}{d\lambda}=0$; then we obtain a generalization of *Poincaré's relative integral invariant*,

$$\int_0^1 d\lambda \int_{S_1(\lambda)}^{(n-1)} P_\alpha \frac{dz^\alpha}{d\lambda} du. \tag{8·4}$$

The P_α and z^α appearing under the integral sign in (8·4) are the initial data which determine the extremal if the boundary data on the time-like part of S are fixed.

The relative integral invariant (8·3) or (8·4) can be transformed into an *absolute integral invariant* with the help of Stokes' theorem. Consider a two-parametric family of extremals, labelled by the parameters ω_1, ω_2 such that the closed one-parametric family forms its boundary. It may be pictured as a "solid" tube of extremals in the (x, z)-space with the given tube as its surface. We then obtain for the integral invariants (8·3) and (8·4) respectively, with an obvious notation,

$$\iint d\omega_1 d\omega_2 \int_{S_1(\omega)}^{(n-1)} \left(\frac{d(X_i, x^i)}{d(\omega_1, \omega_2)} + \frac{d(P_\alpha, z^\alpha)}{d(\omega_1, \omega_2)} \right) du, \tag{8·5}$$

$$\iint d\omega_1 d\omega_2 \int_{S_1(\omega)}^{(n-1)} \frac{d(P_\alpha, z^\alpha)}{d(\omega_1, \omega_2)} du. \tag{8·6}$$

This is the generalization of Poincaré's absolute integral invariant (cp. Cartan 1922, pp. 19–20; Whittaker 1927, p. 272).

Again, without loss of generality we may work with the expression (8·6). As the domain of integration in the parameter space is arbitrary, the integrand under the integral-sign over ω must itself remain invariant under a displacement along the extremals. This integrand is a *generalized Lagrange bracket* (it will be referred to as L.B.) (cp. Whittaker 1927, p. 298),

$$\{\omega_1, \omega_2\} = \int_{S_1}^{(n-1)} \frac{d\{P_\alpha(u), z^\alpha(u)\}}{d(\omega_1, \omega_2)} du. \tag{8·7}$$

In the case $n=1$ the space-like region S_1 consists of a single point so that (8·7) reduces to the ordinary L.B., involving no integration but only the summation over α. In the general case we have the integration over S_1 besides the summation over the number of dependent variables.

In all the subsequent considerations the variables (u^r) will appear on the same footing as the index α. It is therefore useful to consider a point (u) of S_1 as a *continuous index*, which together with the discrete index α serves to label the dependent variables z and their conjugates P. From this standpoint the only difference between the dynamics of point systems and that of continua is the larger *magnitude* of indices which in the latter case is necessary in order to specify the initial data.

This standpoint also approaches the standpoint of Heisenberg and Pauli, who treated the continuum as the limiting case of a very large number of points.

In mechanics, all transformations preserving the L.B.'s are defined as "*canonical transformations*". The same definition will be adopted for the general case. Then the invariance property of the L.B.'s may be expressed by saying that for any dynamical continuum with fixed spatial boundary conditions the motion consists of a continuous unfolding of canonical transformations, starting from some initial data on some space-like region S_1.

9. Poisson brackets

In the considerations of the preceding section we have treated the P_α and z^α as functions of the u^r on S_1 and of the parameters ω_1, ω_2. We shall now invert the connexion and treat the quantities ω_1, ω_2 as functionals of the $P_\alpha(u)$, $z^\alpha(u)$, defined on S_1. The functional differential of such a functional ω has the form

$$\delta\omega = \int_{S_1}^{(n-1)} \left(\frac{\delta\omega}{\delta P_\alpha(u)} \delta P_\alpha(u) + \frac{\delta\omega}{\delta z^\alpha(u)} \delta z^\alpha(u) \right) du, \tag{9·1}$$

where $\dfrac{\delta\omega}{\delta P_\alpha(u)}$ and $\dfrac{\delta\omega}{\delta z^\alpha(u)}$ denote the functional derivatives of ω at the point (u).

In particular, we shall have to consider the functionals $P_\beta(v)$ and $z^\beta(v)$, i.e. the values of P_β and z^β at some given point (v) on S_1. We obtain from (9·1)

$$\left.\begin{aligned} \frac{\delta P_\beta(v)}{\delta P_\alpha(u)} &= \delta^\alpha_\beta \, \delta(v-u), & \frac{\delta P_\beta(v)}{\delta z^\alpha(u)} &= 0, \\ \frac{\delta z^\beta(v)}{\delta P_\alpha(u)} &= 0, & \frac{\delta z^\beta(v)}{\delta z^\alpha(u)} &= \delta^\beta_\alpha \, \delta(v-u), \end{aligned}\right\} \tag{9·2}$$

where $\delta(v-u)$ denotes the product of the $n-1$ δ-functions $\delta(v^r - u^r)$ of Dirac.

For any two functionals ω_1, ω_2 we define the ***generalized Poisson bracket*** (referred to as P.B.) (cp. Whittaker 1927, p. 299) by

$$[\omega_1, \omega_2] = \int_{S_1}^{(n-1)} \frac{d(\omega_1, \omega_2)}{d\{P_\alpha(u), z^\alpha(u)\}} du, \tag{9·3}$$

where the Jacobian under the integral sign is defined as

$$\frac{d(\omega_1,\omega_2)}{d\{P_\alpha(u),z^\alpha(u)\}}=\frac{\delta\omega_1}{\delta P_\alpha(u)}\frac{\delta\omega_2}{\delta z^\alpha(u)}-\frac{\delta\omega_1}{\delta z^\alpha(u)}\frac{\delta\omega_2}{\delta P_\alpha(u)}. \tag{9·4}$$

In the case of mechanics the generalized P.B.'s defined by (9·3) reduce to the ordinary P.B.'s. They also obey the same formal rules as the ordinary P.B.'s.

The P.B.'s have the same invariance properties as the L.B.'s. We shall show this by proving a relationship which allows us to express the P.B.'s in terms of the L.B.'s and vice versa. The prescribing of the initial data $P_\alpha(u)$, $z^\alpha(u)$ on S_1 (always for *fixed* boundary data on the time-like region of S) may be considered as a way of labelling all extremals (with the given fixed boundary data). There is, however, no need to choose especially the initial data for labelling the extremals. Any other set of the same magnitude may be used equally well. Let $\omega_a(v)$ $(a=1,2,\ldots,2\nu;\ v\in S_1)$ be such a set which labels the extremals. We shall then have a one-one-correspondence

$$P_1(u)\ldots P_\nu(u),\quad z^1(u)\ldots z^\nu(u)\leftrightarrow\omega_1(v)\ldots\omega_{2\nu}(v), \tag{9·5}$$

which may be interpreted as a *non-singular functional transformation* between two different modes of labelling the extremals.

Taking any two quantities, $\omega_a(v)$ and $\omega_b(v')$ say, we can form their L.B. and their P.B. The general relationship which we want to prove reads

$$\int_{S_1}^{(n-1)}dv\sum_{a=1}^{a=2\nu}\{\omega_a(v),\omega_b(v')\}\,[\omega_a(v),\omega_c(v'')]=\delta_c^b\,\delta(v'-v'')\quad(b,c=1,2,\ldots,2\nu).^* \tag{9·6}$$

This relation expresses that the matrix formed by all the L.B.'s and the matrix formed by all the P.B.'s are contragredient (i.e. inverse and transposed) to each other. A corresponding relation holds in mechanics, with the only difference that we have here the continuous index v besides the discrete index a which results in the matrices having a continuous range of rows and columns. The proof of (9·6) therefore proceeds on the same lines as the corresponding proof in mechanics (Whittaker 1927, pp. 299–300).

To prove (9·6), we first use the definitions (8·7) and (9·3), (9·4) for the L.B.'s and P.B.'s so that (9·6) becomes

$$\int dv\sum_{a=1}^{a=2\nu}\int du\left(\frac{\delta P_\alpha(u)}{\delta\omega_a(v)}\frac{\delta z^\alpha(u)}{\delta\omega_b(v')}-\frac{\delta z^\alpha(u)}{\delta\omega_a(v)}\frac{\delta P_\alpha(u)}{\delta\omega_b(v')}\right)$$
$$\times\int du'\left(\frac{\delta\omega_a(v)}{\delta P_\beta(u')}\frac{\delta\omega_c(v'')}{\delta z^\beta(u')}-\frac{\delta\omega_a(v)}{\delta z^\beta(u')}\frac{\delta\omega_c(v'')}{\delta P_\beta(u')}\right),$$

---

* It will be seen from the definition (8·7) of the L.B.'s that $\{\omega_a,\omega_b\}$ is contravariant in a and b. Since this fact is not expressed in our notation, we use the Σ-symbol for the summation over a.

it being understood that the integration always extends over the $(n-1)$-dimensional region S_1. In evaluating this expression we use the relations

$$\int dv \sum_{a=1}^{a=2\nu} \frac{\delta P_\alpha(u)}{\delta\omega_a(v)} \frac{\delta\omega_a(v)}{\delta P_\beta(u')} = \delta^\beta_\alpha \delta(u-u'), \quad \int dv \sum_{a=1}^{a=2\nu} \frac{\delta P_\alpha(u)}{\delta\omega_a(v)} \frac{\delta\omega_a(v)}{\delta z^\beta(u')} = 0,$$

$$\int dv \sum_{a=1}^{a=2\nu} \frac{\delta z^\alpha(u)}{\delta\omega_a(v)} \frac{\delta\omega_a(v)}{\delta P_\beta(u')} = 0, \quad \int dv \sum_{a=1}^{a=2\nu} \frac{\delta z^\alpha(u)}{\delta\omega_a(v)} \frac{\delta\omega_a(v)}{\delta z^\beta(u')} = \delta^\alpha_\beta \delta(u-u'),$$

which follow directly from (9·2). We then obtain for the above expression

$$\int du \int du' \left(\frac{\delta z^\alpha(u)}{\delta\omega_b(v')} \frac{\delta\omega_c(v'')}{\delta z^\beta(u')} \delta^\beta_\alpha \delta(u-u') + \frac{\delta P_\alpha(u)}{\delta\omega_b(v')} \frac{\delta\omega_c(v'')}{\delta P_\beta(u')} \delta^\alpha_\beta \delta(u-u') \right)$$
$$= \int du \left(\frac{\delta z^\alpha(u)}{\delta\omega_b(v')} \frac{\delta\omega_c(v'')}{\delta z^\alpha(u)} + \frac{\delta P_\alpha(u)}{\delta\omega_b(v')} \frac{\delta\omega_c(v'')}{\delta P_\alpha(u)} \right) = \delta^b_c \delta(v'-v''),$$

so that (9·6) is proved.

It follows that the P.B.'s are invariant under canonical transformations.

10. Classical laws in terms of Poisson brackets

As in mechanics, we can formulate explicit conditions which are necessary and sufficient for a transformation of the $P_\alpha(u)$, $z^\alpha(u)$ to be canonical. For, from the formulae (9·2)–(9·4), we obtain for trivial reasons

$$\left.\begin{aligned} [P_\alpha(v), P_\beta(v')] = 0, \quad [z^\alpha(v), z^\beta(v')] = 0, \\ [P_\alpha(v), z^\beta(v')] = \delta^\beta_\alpha \delta(v-v'), \end{aligned}\right\} \tag{10·1}$$

and it follows from the invariance of the P.B.'s that the transformed quantities, if obtained by a canonical transformation, must satisfy the same relations. This proves that the relations (10·1) are necessary.

By differentiating the second and the third relation of (10·1) with respect to v'^r we obtain, taking account of (5·1),

$$[z^\alpha(v), \tilde{z}^\beta_r(v')] = 0, \quad [P_\alpha(v), \tilde{z}^\beta_r(v')] = -\delta^\beta_\alpha \delta_r(v-v'), \tag{10·2}$$

where $\delta_r(v-v')$ stands for $\frac{\mathfrak{d}}{\mathfrak{d}u^r}\delta(v-v')$. Now let $\mathscr{F}(P_\alpha, z^\alpha, \tilde{z}^\alpha_r)$ be an arbitrary function defined on S_1 and put

$$\mathsf{F} = \int_{S_1}^{(n-1)} \mathscr{F}(v)\, dv, \quad (v \in S_1). \tag{10·3}$$

We can evaluate the P.B.'s of $P_\alpha(u)$ and $z^\alpha(u)$ with F with the help of (10·1) and (10·2) and obtain

$$[P_\alpha(u), \mathsf{F}] = \frac{\partial\mathscr{F}}{\partial z^\alpha} - \frac{\mathfrak{d}}{\mathfrak{d}u^s} \frac{\partial\mathscr{F}}{\partial\tilde{z}^\alpha_s}, \quad [z^\alpha(u), \mathsf{F}] = -\frac{\partial\mathscr{F}}{\partial P_\alpha}. \tag{10·4}$$

In this manner we can evaluate any P.B.'s, using solely the relations (10·1), the relations (10·2) being a consequence of (10·1). It follows, therefore, that (10·1) is also sufficient for a transformation to be canonical.

We shall now apply the general formulae (10·4) to the quantities $\mathscr{H}$ and h_r, defined in section 5. We put

$$\mathsf{H} = \int_{S_1}^{(n-1)} \mathscr{H}(v)\,dv, \quad \mathsf{h}_r = \int_{S_1}^{(n-1)} h_r(v)\,dv. \tag{10·5}$$

Applying (10·4) to the case $\mathscr{F} = h_r$, we obtain from the last set of (5·3)

$$[P_\alpha(u), \mathsf{h}_r] = \frac{\eth P_\alpha}{\eth u^r}, \quad [z^\alpha(u), \mathsf{h}_r] = \frac{\eth z^\alpha}{\eth u^r}. \tag{10·6}$$

Applying (10·4) to the case $\mathscr{F} = \mathscr{H}$ and using the canonical equations (5·6), we obtain

$$[P_\alpha(u), \mathsf{H}] = \frac{\eth P_\alpha}{\eth w}, \quad [z^\alpha(u), \mathsf{H}] = \frac{\eth z^\alpha}{\eth w}. \tag{10·7}$$

Since S_1 may be any space-like region, w may be any time-like co-ordinate.

Conversely, the canonical equations (5·6) follow from (10·7) with the help of the general formula (10·4). The equations (10·7) may therefore be considered as the formulation of the canonical equations by means of P.B.'s. In the case of mechanics they go over into the well-known formulae

$$[p_\alpha, \mathscr{H}] = \frac{dp_\alpha}{dx}, \quad [z^\alpha, \mathscr{H}] = \frac{dz^\alpha}{dx}. \tag{10·7a}$$

In the case $n > 1$ they are supplemented by the equations (10·6). (10·7) and (10·6) form together a set of covariant vector equations in the x-space.

It follows from (10·6) and (10·7) that we have for an arbitrary function $\mathscr{F}(P_\alpha, z^\alpha, \mathfrak{z}_r^\alpha)$,

$$[\mathscr{F}(u), \mathsf{H}] = \frac{\eth \mathscr{F}}{\eth w}, \quad [\mathscr{F}(u), \mathsf{h}_r] = \frac{\eth \mathscr{F}}{\eth u^r}. \tag{10·8}$$

In this way every dynamical relation of the continuum can be expressed in terms of the P.B.'s.

11. QUANTIZATION

Having formulated classical continuum dynamics in terms of Poisson brackets, the quantization can be performed in the same way as in mechanics. The continuous set of initial data $P_\alpha(u)$, $z^\alpha(u)$ becomes a continuous set of non-commuting observables, and a general argument, which is valid irrespective of whether the set is discrete or continuous, requires that the quantum P.B.'s, which are defined by the property of obeying the same formal rules as the classical P.B.'s, must be proportional to the commutation brackets. In this way we obtain ultimately (Dirac 1935, pp. 89–90)

$$[\omega_1, \omega_2] = \frac{i}{\hbar}(\omega_1\omega_2 - \omega_2\omega_1). \tag{11·1}$$

With this new interpretation of the P.B.'s the relations (10·1) and 10·2) become the "*Quantum Conditions*", the function $\mathscr{H}(P_\alpha, z^\alpha, \tilde{z}^\alpha_\tau)$ or better its integral H becomes the "*Quantum Hamiltonian*", and the relations (10·7) become the "*Quantum Equations of Motion*" of the quantized continuum. The Hamiltonian has an invariant significance, as we have seen at the end of section 5. The relations (10·4) and (10·6) follow from (10·1) and (10·2), and (10·8) follows from (10·6) and (10·7) by the same arguments as in the classical theory.

The quantum formalism thus obtained is the same as the one we had obtained in I (Weiss 1936, pp. 201–4) by an argument based on a comparison with the results of Heisenberg and Pauli. The present formalism applies to any space-like region S_1 and the corresponding time-like co-ordinate w normal to S_1, but it does not and, on physical grounds, should not apply to a time-like region.*

SUMMARY

The Hamilton-Jacobi theory of point mechanics is extended to the mechanics of continuous media, following on the lines first proposed by Prange. The method is based on the equivalence between the Euler equations and the "boundary formula" (*formule aux limites*) in the calculus of variations. It is shown that the notions of Lagrange brackets and Poisson brackets can be extended, but only if the Euler equations of the continuum are of hyperbolic type. Consequently, these notions only apply to a dynamical continuum and not to equilibrium problems. Except for this restriction, the method is applicable quite generally, for linear as well as for non-linear theories. Once the dynamical laws are expressed in terms of Poisson brackets, the transition to the quantum theory can be effected by a brief, formal argument in the same way as in point mechanics.

---

* At the end of I, I tried to apply the quantization to a time-like region as well. This was due to my not having sufficiently taken account of the hyperbolic character of all dynamical problems. No quantum relations on a time-like region are justified.

REFERENCES

Born 1934 *Proc. Roy. Soc.* A, **143**, 410.
Cartan 1922 "Leçons sur les invariants intégraux." Paris.
Courant-Hilbert 1937 "Methoden der mathematischen Physik", **2**. Berlin.
De Donder 1935 "Théorie invariantive du calcul des variations." Paris.
Dirac 1935 "The principles of quantum mechanics", 2nd ed. Oxford.
Hadamard 1910 "Leçons sur le calcul des variations." Paris.
— 1923 "Lectures on Cauchy's problem." Yale, Oxford.
— 1932 "Leçons sur le problème de Cauchy etc." Paris.
Heisenberg and Pauli 1929 *Z. Phys.* **56**, 1.
Hilbert 1900 *Nachr. Ges. Wiss. Göttingen*, p. 253.
— 1905 *Nachr. Ges. Wiss. Göttingen*, p. 159.
Juvet 1926 "Sur une équation aux dérivées fonctionnelles partielles etc." Thèse, Paris.
Prange 1915 "Die Hamilton-Jacobi'sche Theorie für Doppelintegrale." Dissertation, Göttingen.
Volterra 1913 "Leçons sur les fonctions de lignes." Paris.
— 1930 "Theory of functionals." London.
Volterra and Pérès 1937 "Théorie générale des fonctionnelles", **1**. Paris.
Weiss 1936 *Proc. Roy. Soc.* A, **156**, 192.
Whittaker 1927 "Analytical dynamics", 3rd ed. Cambridge.

On the Hamilton-Jacobi theory and quantization of generalized electrodynamics

BY P. WEISS, PH.D., *Downing College, Cambridge*

(*Communicated by R. H. Fowler, F.R.S.—Received* 1 *September* 1938)

1. INTRODUCTION

In the preceding paper, which will be quoted as A, the Hamilton-Jacobi theory has been developed for a dynamical continuum of quite general type, and it has been shown that its classical laws can be expressed in terms of Poisson brackets involving pairs of canonically conjugate variables. In this way the theory of a dynamical continuum can be treated on the same lines as the dynamical theory of point systems.

This result is of importance for the study of the procedure of quantization, for it is a well-known fact that a quantum Poisson bracket which obeys the same algebraic rules as a classical Poisson bracket must be proportional to a commutation bracket (Dirac 1935, p. 89). This fact is independent of whether the observables in question belong to a finite or infinite or even continuous set of observables, and it applies therefore to a continuum theory as well as to point mechanics.

Thus one has an extension to continuum physics of the satisfactory feature of point mechanics that the laws of the quantum theory, though differing fundamentally from the laws of the classical theory in *interpretation*, can be given the same *form* as these.

The purpose of the present paper is to apply this procedure to the theory of the electromagnetic field. We shall not only deal with Maxwell's field but also with the more general field of Born (1934, 1937; Born and Infeld 1934*a*), of which Maxwell's field is the linear case. We shall also deal with the still more general field of Mie (1912*a*, *b*, 1913), which involves the potentials explicitly, while Born's field involves them only through their derivatives. The linear case of Mie's field has recently been studied by Proca (1936) and has then been taken by Bhabha (1938) and Kemmer (1938) to describe the wave-field of the heavy electron. The linear cases are probably more important for physics than the non-linear ones, but for the procedure of the Hamilton-Jacobi theory it does not make any difference whether one deals with a linear or with a non-linear case so that we can treat the general case without incurring additional difficulties.

All these cases of generalized electrodynamics have certain peculiar features in common, which are mainly rooted in the fact that the derivatives of the potentials occur only in their antisymmetrical combinations (field strengths). We shall refer to this fact as the "*antisymmetry*" of electrodynamics. The potentials form a vector in space-time, and the antisymmetry can be accounted for by the postulate of general relativistic invariance, independent of the affine connexion of space-time. In Born's case it can instead be accounted for by the postulate of gauge-invariance.†

This antisymmetry gives rise to well-known anomalies concerning the quantum laws of electrodynamics (Heisenberg and Pauli 1929), and it is mainly for their sake that a special discussion of the electromagnetic field, beyond the considerations of A, is necessary.

We wish to state the result of this discussion at the outset: *The considerations of A show that, in order to express the classical laws in terms of Poisson brackets, it is first of all necessary to describe the system in terms of an independent and complete set of initial data so that, for given boundary data in physical space, the solution in space-time is completely determined by these initial data and depends on them continuously.*

---

† For a more detailed discussion of the special features of electrodynamics cf. Weiss (1936). Some considerations of that paper will be repeated here.

In the general theory of A, the dependent variables together with their conjugates form such an independent and complete set of initial data. But in the case of electrodynamics this ceases to be true. The antisymmetry imposes a restriction on this set of initial data, so that a new set of independent data has to be formed. This new set will be different according to whether we use Mie's or Born's electrodynamics, as we shall see later on. But, whatever the independent and complete set may be in each particular case, *once we have obtained it, the general procedure of A applies, and, using it, we can formulate the classical laws in terms of Poisson brackets and therefore the quantum laws without arriving at contradictions.*

Another feature of all electrodynamics is its relativistic invariance. Now, our procedure, if applied to an invariant theory, is itself of invariant form. This is due to the fact, discussed in A, § 6, that the initial data are given on an arbitrary space-like section S_1 which may but need not coincide with a section of constant time in a given co-ordinate system. All relations in the Hamilton-Jacobi theory as well as in the quantum theory of the continuum refer to the *parameters* u^r of the section S_1 and to a co-ordinate w perpendicular to S_1. They are therefore independent of the co-ordinate system in space-time. *No explicit proof of the invariance of the quantum conditions of electrodynamics is therefore necessary according to our procedure.*†

Furthermore, for a given space-like S_1, the co-ordinate w has an invariant significance, and so has therefore its conjugate which, as we have seen in A, plays the role of Hamiltonian in continuum dynamics. In other words: *The Hamiltonian is invariantly connected with the initial section* S_1. If and only if we use a co-ordinate system, in which "space" for a given time coincides with S_1, then does the Hamiltonian coincide with the time-component of the stress-energy-momentum tensor. This explains the fundamental role played by that time-component in the customary procedures of quantizing fields. For in all these procedures, S_1 is from the beginning identified with a section of constant time.

The relativistic invariance gives rise to an ambiguity in the definition of the stress-energy-momentum tensor. It may be defined by the boundary formula according to the procedure of A and in agreement with mechanics and is denoted by U^i_k. It may also be defined by varying the Lagrangian with respect to the metric tensor and is then denoted by T^i_k.

For gauge-invariant electrodynamics, i.e. for the Born-Maxwell case, we shall show that the two definitions of the stress-energy-momentum tensor coincide provided that we develop the Hamilton-Jacobi theory in really gauge-invariant form. This will necessitate a slight modification of the procedure of A.

---

† Explicit proofs have been given by Heisenberg and Pauli (1929) for Maxwell's case and by Infeld (1937) and Pryce (1937) for Born's case.

In Mie's electrodynamics, however—and therefore also in Proca's case—which is not gauge-invariant, the ambiguity subsists, but we shall show that both tensors lead to the same physical results.

We shall treat the Mie-Proca case in §§ 2 and 3 and the Born-Maxwell case in the subsequent sections. The resulting scheme of quantum electrodynamics will be the customary one (Heisenberg and Pauli 1929; Born and Infeld 1934*b*, 1935); the difference which, we think, justifies the present considerations lies in the *procedure* leading up to the quantization.

2. MIE'S ELECTRODYNAMICS

The action integral reads

$$I = \int_D^{(4)} \mathscr{L}(g_{ik}, \phi_i, f_{ik})\,dx \quad (i, k, l = 0, 1, 2, 3). \tag{2·1}$$

Here the tensor $g_{ik} = g_{ki}$ denotes the metric; we assume space-time to be flat and choose a co-ordinate system in which the g_{ik} are constant and $\det|g_{ik}| = -1$. We need, therefore, not distinguish between tensors and tensor densities.

The covariant vector ϕ_i denotes the electromagnetic potentials, and the covariant tensor $f_{ik} = -f_{ki}$, defined by

$$f_{ik} = \frac{\partial \phi_k}{\partial x^i} - \frac{\partial \phi_i}{\partial x^k}, \tag{2·2}$$

denotes the field strengths. D is a domain in space-time, whose boundary is called S. The partial derivatives of the Lagrangian $\mathscr{L}$ are defined by the formula

$$d\mathscr{L} = \tfrac{1}{2}T^{ik}dg_{ik} + \rho^i d\phi_i + \tfrac{1}{2}p^{ik}df_{ik}. \tag{2·3}$$

The field equations consist of the integrability conditions

$$\frac{\partial f_{kl}}{\partial x^i} + \frac{\partial f_{li}}{\partial x^k} + \frac{\partial f_{ik}}{\partial x^l} = 0, \tag{2·4}$$

and of the Euler equations

$$\frac{\partial p^{lk}}{\partial x^l} = \rho^k. \tag{2·5}$$

Applying the procedure of A, without varying the metric, we obtain as *boundary formula* (cf. A, (3·5), (3·6), (4·1))

$$\delta I = \oint_\omega^{(3)} (X_k \delta x^k + P^i \delta \phi_i)\,du, \tag{2·6}$$

where we have put
$$X_k = U^l_k N_l, \quad P^i = p^{li} N_l, \tag{2·7}$$
$$U^l_k = \mathscr{L}\delta^l_k - p^{lj}\frac{\partial \phi_j}{\partial x^k} \tag{2·8}$$
(cf. A, (3·6)). The N_l in (2·7) denote a vector normal to S.

The pairs of conjugate variables are
$$(X_k, x^k); \quad (P^i, \phi_i). \tag{2·9}$$
Owing to the antisymmetry, $p^{ik} = -p^{ik}$, we have
$$P^i N_i = 0, \tag{2·10}$$
so that the P^i are not independent, the vector $\{P^i\}$ always being tangential to S.

The tensors defined in (2·3) are connected by the relation
$$T^{il} g_{kl} + \rho^i \phi_k + p^{il} f_{kl} = \mathscr{L}\delta^i_k, \tag{2·11}$$
which is a consequence of the relativistic invariance (Born 1937, p. 22). With the notation
$$T^{il} g_{kl} = T^i_k, \tag{2·12}$$
we obtain from (2·8) and (2·11)
$$T^i_k = U^i_k + p^{ij}\frac{\partial \phi_k}{\partial x^j} - \rho^i \phi_k,$$
or, using (2·5) and the antisymmetry of the p^{ij},
$$T^i_k = U^i_k + \frac{\partial}{\partial x^j}(p^{ij}\phi_k). \tag{2·13}$$
Thus the tensors T^i_k and U^i_k only differ by a divergence. They both satisfy the conservation law
$$\frac{\partial T^l_k}{\partial x^l} = 0, \quad \frac{\partial U^l_k}{\partial x^l} = 0. \tag{2·14}$$

Let us now consider a space-like section S_1, referred to parameters u^r ($r = 1, 2, 3$). We wish to find an independent and complete set of initial data on S_1. For this purpose we introduce a "*natural co-ordinate system*" (cf. A, § 5). We then obtain
$$N_0 = 1, \quad N_1 = N_2 = N_3 = 0,$$
and therefore
$$X_k = U^0_k; \quad P^0 = 0, \quad P^r = p^{0r}. \tag{2·15}$$
Of the Euler equations (2·5) the one with $k = 0$ becomes
$$\frac{\partial p^{r0}}{\partial u^r} = -\frac{\partial P^r}{\partial u^r} = \rho^0. \tag{2·16}$$

ρ^0 depends, among others, on the potentials and we can, for instance, resolve (2·16) with respect to ϕ_0. Thus we see that $P^r(u)$, $\phi_r(u)$ may be taken as an independent set of data on S_1, while $P^0(u)$ and $\phi_0(u)$ are then determined by (2·15) and (2·16). Our set is thus complete.

The considerations of A apply word for word to this new set of data and we obtain for their P.B.'s from A, eq. (10·1),

$$\left.\begin{aligned} &[P^r(u), P^s(u')] = 0, \quad [\phi_r(u), \phi_s(u')] = 0 \\ &[P^r(u), \phi_s(u')] = \delta^r_s\, \delta(u-u'). \end{aligned}\right\} \quad (r, s = 1, 2, 3) \qquad (2{\cdot}17)$$

From (2·16) we then obtain

$$[\rho^0(u), P^r(u')] = 0, \quad [\rho^0(u), \phi_r(u')] = -\delta_r(u-u'). \qquad (2{\cdot}18)$$

(2·17) and (2·18) together define implicitly the P.B.'s of ϕ_0 with P^r and ϕ_r.

Let us define, in a natural co-ordinate system (w, u^r),

$$\mathscr{H}(u) = T^0_0 = \mathscr{L} - \rho^0\phi_0 - P^s f_{0s}, \quad h_r(u) = T^0_r = -\rho^0\phi_r - P^s \bar{f}_{rs}, \qquad (2{\cdot}19)$$

where we have
$$f_{0s} = \frac{\partial\phi_s}{\partial w} - \frac{\partial\phi_0}{\partial u^s}, \quad \bar{f}_{rs} = \frac{\partial\phi_s}{\partial u^r} - \frac{\partial\phi_r}{\partial u^s}, \qquad (2{\cdot}2a)$$

and put
$$\mathsf{H} = \int_{S_1}^{(3)} \mathscr{H}(v)\, dv, \quad \mathsf{h}_r = \int_{S_1}^{(3)} h_r(v)\, dv. \qquad (2{\cdot}20)$$

From (2·13) we obtain

$$\mathscr{H} = U^0_0 + \frac{\partial}{\partial u^s}(P^s\phi_0), \quad h_r = U^0_r + \frac{\partial}{\partial u^s}(P^s\phi_r). \qquad (2{\cdot}21)$$

Therefore, H and h_r differ from the corresponding quantities in A, defined by means of the tensor U^i_k, only by expressions referring to the boundary of S_1, and for their P.B.'s with any quantity taken at the interior of S_1 it is thus immaterial whether we consider T^i_k or U^i_k as the stress-energy-momentum tensor. The former has the advantage of being expressible in terms of the field strengths and it is therefore usually used in physics.

The Hamiltonian $\mathscr{H}$ given by the first equation of (2·19) can, after performing a Legendre transformation (A, §5)

$$\phi_0 \to \rho^0, \quad f_{0r} \to P^r, \quad \mathscr{L} \to \mathscr{H}$$

and using (2·16), be expressed as

$$\mathscr{H} = \mathscr{H}\left(\frac{\partial P^s}{\partial u^s}, \phi_r, P^r, \bar{f}_{rs}\right). \qquad (2{\cdot}22)$$

Taking account of (2·2a), one sees that the Hamiltonian $\mathscr{H}$ depends only on the initial data $\phi_r(u)$, $P^r(u)$ on S_1 as it should do. The differential of $\mathscr{H}$ reads

$$d\mathscr{H} = -\phi_0 d\rho^0 + \rho^r d\phi_r - f_{0r} dP^r + \tfrac{1}{2} p^{rs} d\bar{f}_{rs}. \qquad (2{\cdot}23)$$

Let us work out the P.B.'s involving H and h_r. We obtain from (2·16)–(2·19) and from (2·23)

$$[\phi_r(u), \mathsf{H}] = \frac{\partial \phi_0}{\partial u^r} + f_{0r} = \frac{\partial \phi_r}{\partial u^0}, \quad [\phi_r(u), \mathsf{h}_s] = \frac{\partial \phi_s}{\partial u^r} + f_{sr} = \frac{\partial \phi_r}{\partial u^s}. \tag{2·24}$$

Similarly $$[P^r(u), \mathsf{H}] = \rho^r - \frac{\partial p^{sr}}{\partial u^s}, \quad [P^r(u), \mathsf{h}_s] = \frac{\partial P^r}{\partial u^s}. \tag{2·25}$$

From the last set together with (2·5) it follows that

$$[P^r(u), \mathsf{H}] = \frac{\partial P^r}{\partial w}. \tag{2·26}$$

The relations (2·17), (2·22), (2·26) represent the classical laws of Mie's electrodynamics in terms of P.B.'s, and the quantum laws are formally the same.

3. THE PROCA FIELD

The case of Proca, written in relativistic notation, is described by the particular Lagrangian

$$\mathscr{L} = \tfrac{1}{4} f_{ik} \bar{f}^{ik} + \tfrac{1}{2}\lambda^2 \phi_i \bar{\phi}^i. \tag{3·1}$$

The field quantities are complex. A bar denotes the conjugate complex quantity. The contravariant tensors f^{ik}, ϕ^i are defined as

$$f^{ik} = g^{il} g^{km} f_{lm}, \quad \phi^i = g^{il} \phi_l. \tag{3·2}$$

λ is a real constant. From (2·3) we obtain in this case

$$\rho^i = \lambda^2 \bar{\phi}^i, \quad p^{ik} = \bar{f}^{ik} \tag{3·3}$$

and the conjugate complex relations. The Euler equations (2·4) go over into

$$\frac{\partial f^{lk}}{\partial x^l} = \lambda^2 \phi^k, \tag{3·4}$$

and into the conjugate complex relations. From the second set of (2·15) we see that $f^{0r}(u)$ and $\phi_r(u)$ form a complete and independent set of initial data, while f_{rs} and ϕ_0 are derived from the relations (2·2*a*) and from (2·16) which reads here

$$\phi_0 = -\frac{1}{\lambda^2} \frac{\partial f^{0r}}{\partial u^r}. \tag{3·5}$$

The P.B.'s (2·17) and (2·18) simplify to

$$\left.\begin{aligned}&[f^{0r}(u), f^{0s}(u')] = 0, \quad [\phi_r(u), \phi_s(u')] = 0,\\ &[\bar{f}^{0r}(u), \phi_s(u')] = \delta^r_s \delta(u-u'),\\ &[\phi^0(u), f^{0r}(u')] = 0, \quad [\bar{\phi}^0(u), \phi_r(u')] = -\frac{1}{\lambda^2}\delta_r(u-u'),\end{aligned}\right\} \qquad (3·6)$$

and to the conjugate complex relations. Thus in Proca's case the P.B.'s for the scalar potential ϕ_0 are given *explicitly*. The corresponding quantum commutation rules are formally the same (Kemmer 1938, eq. (12)). In this way we obtain the quantized Proca field in relativistically invariant form.

4. CONSEQUENCES OF THE POSTULATE OF GAUGE-INVARIANCE

We have pointed out in the introduction that Born's and in particular Maxwell's electrodynamics possess the property of *gauge-invariance* and that this fact necessitates a modification of the procedure of A. This modification we now wish to study.

By gauge-invariance one understands that no physical change is involved when the potentials ϕ_i are replaced by $\phi_i + \partial\chi/\partial x^i$, χ being an arbitrary scalar. Only quantities which remain invariant under such a "gauge-transformation"

$$\phi_i \to \phi_i + \frac{\partial \chi}{\partial x^i} \qquad (4·1)$$

have a physical meaning. The ϕ_i themselves have therefore no physical meaning but the functional, more precisely, the "line-function"

$$\phi = \phi[C] = \oint_C \phi_i dx^i \qquad (4·2)$$

has. Here C represents any closed (one-dimensional) curve in space-time, and ϕ is a functional of C.

It should be noted that in four-dimensional space-time the curves C are in no way affected by the presence of singular lines representing charges, since they can always be contracted to a point without meeting the singular lines. Their topological role in four-space is analogous to the topological role of a pair of points in three-space.

The line-function $\phi[C]$, *given by* (4·2), *expresses what one really means by "potential" in all relativistic and gauge-invariant electrodynamics.*

The dependent quantity now consists not of a set of functions, like the z^α in A, but of the line-function $\phi[C]$. Its derivatives, however, the so-called

"Grassmann-derivatives", are ordinary functions, viz. the field strengths f_{ik}, defined by

$$f_{ik} = \lim_{\Delta S^{ik}\to 0} \frac{\oint_C \phi_i dx^i}{\Delta S^{ik}}, \tag{4·3}$$

where ΔS^{ik} is the area of a two-dimensional surface element parallel to the (x^i, x^k)-plane (more precisely: *asymptotically* parallel, for $\Delta S^{ik} \to 0$) and C is the boundary curve of ΔS^{ik}. The definition (4·3) is equivalent with (2·2); this fact is the essence of Stokes's theorem in four-dimensional space.

We now have to determine the variations of x^i and $\phi[C]$ corresponding to those of A, § 3. We define a variational line-function $\eta[C]$, analogously to A, (3·1), by

$$\bar\phi[C] = \phi[C] + \epsilon\eta[C], \tag{4·4}$$

and variations δx^i, $\delta\phi[C]$, analogously to A, (3·2), by

$$\bar x^i = x^i + \epsilon\delta x^i, \quad \bar\phi[\bar C] = \phi[C] + \epsilon\delta\phi[C]. \tag{4·5}$$

According to (4·2), the equation (4·4) and the second part of (4·5) read

$$\oint_C \bar\phi_i(x)\,dx^i = \oint_C \phi_i(x)\,dx^i + \epsilon\oint_C \eta_i(x)\,dx^i \tag{4·6}$$

and

$$\oint_{\bar C} \bar\phi_i(\bar x)\,d\bar x^i = \oint_C \phi_i(x)\,dx^i + \epsilon\oint_C \delta\phi_i(x)\,dx^i \tag{4·7}$$

respectively.

Let us evaluate the left-hand side of (4·7). We obtain, neglecting powers of ϵ higher than the first,

$$\oint_{\bar C} \bar\phi_i(\bar x)\,d\bar x^i = \oint_{\bar C} \phi_i(\bar x)\,d\bar x^i + \epsilon\oint_C \eta_i(x)\,dx^i. \tag{4·8}$$

And from the first set of (4·5) we get

$$\begin{aligned}\oint_{\bar C} \phi_i(\bar x)\,d\bar x^i &= \oint_C \phi\,\delta_i(x+\epsilon\delta x)\,(dx^i + \epsilon d\,\delta x^i)\\ &= \oint_C \phi_i(x)\,dx^i + \epsilon\oint_C\left(\frac{\partial\phi_i}{\partial x^k}\delta x^k + \phi_k\frac{\partial\delta x^k}{\partial x^i}\right)dx^i,\end{aligned}$$

where in the last term under the second integral sign we have interchanged the summation indices i and k. Performing an integration by parts on this term we get, taking into account that C is closed and using (2·2),

$$\oint_{\bar C} \phi_i(\bar x)\,d\bar x^i = \oint_C \phi_i(x)\,dx^i + \epsilon\oint_C f_{ki}(x)\,\delta x^k dx^i. \tag{4·9}$$

Inserting (4·9) into (4·8), we obtain

$$\oint_{\bar{C}} \bar{\phi}_i(\bar{x})\, d\bar{x}^i = \oint_C \phi_i(x)\, dx^i + \epsilon \oint_C \{f_{ki}(x)\, \delta x^k + \eta_i(x)\}\, dx^i. \tag{4·10}$$

Comparing (4·10) with (4·7), we see that we have

$$\delta\phi_i \sim f_{ki}\, \delta x^k + \eta_i, \tag{4·11}$$

the equivalence holding in the sense of gauge-invariance, and from (4·4) and (4·5) we have

$$\delta\phi[C] = \oint_C f_{ki}\, \delta x^k dx^i + \eta[C]. \tag{4·12}$$

The procedure of A would have given the non-gauge-invariant formula

$$\delta\phi_i = \frac{\partial \phi_i}{\partial x^k} \delta x^k + \eta_i$$

(cf. A, (3·3)) instead of (4·11) or (4·12).

5. BORN'S ELECTRODYNAMICS

The postulate of gauge-invariance, if applied to Mie's electrodynamics, gives the restriction

$$\rho^i = 0, \tag{5·1}$$

so that the action integral becomes

$$\mathsf{I} = \int_D^{(4)} \mathscr{L}(g_{ik}, f_{ik})\, dx. \tag{5·2}$$

The formula (2·3) reduces to

$$d\mathscr{L} = \tfrac{1}{2} T^{ik} dg_{ik} + \tfrac{1}{2} p^{ik} df_{ik}. \tag{5·3}$$

Instead of (2·5) we have the field equations

$$\frac{\partial p^{ik}}{\partial x^i} = 0, \tag{5·4}$$

and instead of (2·11) we now have

$$T^{il} g_{kl} + p^{il} f_{kl} = \mathscr{L} \delta^i_k, \tag{5·5}$$

or, using (2·12),

$$T^i_k = \mathscr{L} \delta^i_k - p^{il} f_{kl}. \tag{5·6}$$

In particular Maxwell's case is described by the Lagrangian (compare with (3·1), (3·2))

$$\mathscr{L} = \tfrac{1}{4} f_{ik} f^{ik} \quad \text{so that} \quad p^{ik} = f^{ik}.$$

If we now calculate the boundary formula, but use the modified procedure of the preceding section, we obtain

$$\delta\mathsf{I} = \oint_S^{(3)} (X_k \delta x^k + P^i \delta\phi_i)\, du, \tag{5·7}$$

where we have now put

$$X_k = T^l_k N_l, \quad P^i = p^{li} N_l. \tag{5·8}$$

This should be compared with (2·7): *The appearance of the T^l_k instead of the U^l_k in the definition of the X_k is due to the gauge-invariant procedure of the preceding section.* The P^i are unchanged.

The gauge-invariance of the boundary formula (5·7) is not obvious on account of the terms $P^i \delta\phi_i$, and we therefore wish to verify it. It will be sufficient to show that $\int_S^{(3)} P^i \frac{\partial\chi}{\partial x^i} du$ vanishes for an arbitrary function $\chi(x)$. We have from (5·8), using Green's theorem,

$$\oint_S^{(3)} P^i \frac{\partial\chi}{\partial x^i} du = \oint_S^{(3)} p^{li} \frac{\partial\chi}{\partial x^i} N_l du = \int_D^{(4)} \frac{\partial}{\partial x^l}\left(p^{li} \frac{\partial\chi}{\partial x^i}\right) dx$$
$$= \int_D^{(4)} \left(\frac{\partial p^{li}}{\partial x^l} \frac{\partial\chi}{\partial x^i} + p^{li} \frac{\partial^2\chi}{\partial x^l \partial x^i}\right) dx.$$

But the integrand under the last integral sign vanishes for an arbitrary $\chi(x)$ on account of the field equations (5·4) and of the antisymmetry of the p^{li}.

We can also calculate the boundary formula for the *dual* action integral (Born and Infeld 1934*a*, p. 435; Weiss 1936, p. 211)

$$\mathsf{I}^* = \int_D^{(4)} \mathscr{L}^*(g_{ik}, p^*_{ik})\, dx.\dagger \tag{5·2 a}$$

The differential of $\mathscr{L}^*$ becomes

$$d\mathscr{L}^* = \tfrac{1}{2} T^{ik} dg_{ik} + \tfrac{1}{2} f^{*ik} dp^*_{ik}. \tag{5·3a}$$

The dual theory is gauge-invariant in the antipotentials ψ^*_i, and the boundary formula becomes

$$\delta\mathsf{I}^* = \oint_S^{(3)} (X_k \delta x^k + F^{*i} \delta\psi^*_i)\, du, \tag{5·7a}$$

where we have put

$$X_k = T^l_k N_l, \quad F^{*i} = f^{*li} N_l. \tag{5·8a}$$

---

† Usually $\mathscr{L}^*$ has been denoted by $\mathscr{H}$ and has been called "Hamiltonian", owing to the formal resemblance of its definition

$$\mathscr{L}^* = \mathscr{L} - \tfrac{1}{2} p^{lm} f_{lm}$$

to that of the Hamiltonian of mechanics. In fact, $\mathscr{L}^*$ does not play the role of Hamiltonian but that of dual Lagrangian. We therefore prefer to reserve the letter $\mathscr{H}$ for the true Hamiltonian, defined, for example, by (5·9) below.

It will be seen that the quantities X_k, which are conjugate to the x^k, are the same in both dual theories. This would not have been the case if we had used the non-gauge-invariant procedure of A.

We now have to look for a complete, independent and *gauge-invariant* set of initial data on S_1. The potentials ϕ_i (or antipotentials ψ_i^*) are not admissible now, since they are not gauge-invariant. The field strengths f_{ik} (or p_{ik}^*), however, are admissible, for they are gauge-invariant.

Introducing a natural co-ordinate system we have, analogously to (2·15), (2·19),

$$\begin{aligned} X_0 = T_0^0 = \mathscr{L} - P^s f_{0s} &= \mathscr{L}^* - F^{*s} p_{0s}^*; \quad X_r = T_r^0 = -P^s \tilde{f}_{rs} = -F^{*s} \tilde{p}_{rs}^* \qquad (5{\cdot}9) \\ &= \mathscr{H} \qquad\qquad\qquad\qquad\qquad\qquad\qquad = h_r \end{aligned}$$

and
$$P^0 = 0, \quad P^r = p^{0r}, \qquad (5{\cdot}10)$$

$$F^{*0} = 0, \quad F^{*r} = f^{*0r}. \qquad (5{\cdot}10a)$$

From the two sets of field equations, which may be written as

$$\frac{\partial p^{lk}}{\partial x^l} = 0, \quad \frac{\partial f^{*lk}}{\partial x^l} = 0, \qquad (5{\cdot}11)$$

we obtain for $k = 0$ restrictions on the initial data, viz.

$$\frac{\partial p^{0r}}{\partial u^r} = \frac{\partial P^r}{\partial u^r} = 0, \qquad (5{\cdot}12)$$

$$\frac{\partial f^{*0r}}{\partial u^r} = \frac{\partial F^{*r}}{\partial u^r} = 0. \qquad (5{\cdot}12a)$$

Subject to these conditions, the six quantities P^r, F^{*r} form an independent and gauge-invariant set of initial data. We wish to show that this set is also complete. For this purpose we observe that we have

$$(P^1, P^2, P^3) = (\tilde{p}_{23}^*, \tilde{p}_{31}^*, \tilde{p}_{12}^*), \qquad (5{\cdot}13)$$

$$(F^{*1}, F^{*2}, F^{*3}) = (-\tilde{f}_{23}, -\tilde{f}_{31}, -\tilde{f}_{12}), \qquad (5{\cdot}13a)$$

and perform the Legendre transformation

$$f_{0r} \rightarrow P^r, \quad \mathscr{L} \rightarrow \mathscr{H}, \qquad (5{\cdot}14)$$

or the dual one
$$p_{0r}^* \rightarrow F^{*r}, \quad \mathscr{L}^* \rightarrow \mathscr{H}. \qquad (5{\cdot}14a)$$

In either case we obtain $\mathscr{H}$ as a function of P^r and F^{*r}, and the derivatives of $\mathscr{H}$ are given by

$$d\mathscr{H} = -f_{0r}\, dP^r - p_{0r}^*\, dF^{*r}. \qquad (5{\cdot}15)$$

Thus the remaining field strengths are determined by P^r, F^{*r}. It should be noted that the present formalism is quite symmetrical between the two dual theories.

Working out the P.B.'s we obtain

$$\left.\begin{aligned}[][P^r(u), P^s(u')] = 0, \quad [\tilde{f}_{qr}(u), \tilde{f}_{st}(u')] = 0,\\ [P^r(u), \tilde{f}_{st}(u')] = \delta^r_s\,\delta_t(u-u') - \delta^r_t\,\delta_s(u-u'),\end{aligned}\right\} \qquad (5\cdot 16)$$

or, in dual formulation,

$$\left.\begin{aligned}[][F^{*r}(u), F^{*s}(u')] = 0, \quad [\tilde{p}^*_{qr}(u), \tilde{p}^*_{st}(u')] = 0,\\ [F^{*r}(u), \tilde{p}^*_{st}(u')] = \delta^r_s\,\delta_t(u-u') - \delta^r_t\,\delta_s(u-u').\end{aligned}\right\} \qquad (5\cdot 16a)$$

The relations $(5\cdot 16a)$ are in substance the same as $(5\cdot 16)$.

With the help of

$$\mathsf{H} = \int_{S_1}^{(3)} \mathscr{H}(v)\,dv, \quad \mathsf{h}_r = \int_{S_1}^{(3)} h_r(v)\,dv \qquad (2\cdot 20)$$

and of the field equations, we obtain

$$[P^r(u), \mathsf{H}] = \frac{\partial P^r}{\partial w}, \qquad [P^r(u), \mathsf{h}_s] = \frac{\partial P^r}{\partial u^s}, \qquad (5\cdot 17)$$

$$[F^{*r}(u), \mathsf{H}] = \frac{\partial F^{*r}}{\partial w}, \quad [F^{*r}(u), \mathsf{h}_s] = \frac{\partial F^{*r}}{\partial u^s}, \qquad (5\cdot 18)$$

and hence for an arbitrary function $\mathscr{G}(P^r, F^{*r})$,

$$[\mathscr{G}(u), \mathsf{H}] = \frac{\partial \mathscr{G}}{\partial w}, \quad [\mathscr{G}(u), \mathsf{h}_s] = \frac{\partial \mathscr{G}}{\partial u^s}. \qquad (5\cdot 19)$$

The quantum laws are, as always, formally the same as the equations $(5\cdot 16)$–$(5\cdot 19)$.

If the space-like section S_1 is especially a section of constant time, we obtain, using space-vector notation,

$$(u^r) = (x, y, z); \quad w = t, \qquad (5\cdot 20)$$

$$(P^r) = -\mathbf{D}; \quad (F^{*r}) = -\mathbf{B}, \qquad (5\cdot 21)$$

$$(h_r) = (\mathbf{D}\times\mathbf{B}) = \mathbf{S}; \quad \mathscr{H} = T^0_0 = U(\mathbf{D}, \mathbf{B}). \qquad (5\cdot 22)$$

The relations $(5\cdot 16)$–$(5\cdot 19)$ go therewith over into the quantum relations of Born and Infeld (1934*b*)—and in Maxwell's case into those of Heisenberg-Pauli. Our formalism thus constitutes an invariant form of quantum electrodynamics.†

---

† In a previous paper (Weiss 1936) an argument more formal than the present one suggested the same quantum relations as obtained here on S_1, and in addition to them other quantum relations on a time-like section. The latter ones do not correspond to initial data and have therefore no physical significance according to the procedure of A.

6. CONCLUDING REMARKS

Our procedure of obtaining Poisson and quantum commutation brackets is not restricted to the preceding cases, where the dependent variables form a vector (or a line-function formed by means of a vector). The procedure applies equally well to cases where the dependent variable is a scalar (Pauli and Weisskopf 1934) or a tensor of higher rank; moreover, any kind of invariant auxiliary conditions may be introduced.

For all such theories one has to find an independent and complete set of initial data. Then the description of the classical laws by means of Poisson brackets and of the quantum laws by means of commutation brackets is determined.

One might think of applying the same procedure to the quantization of electron waves which obey the Dirac equations. Since these equations are of the first order, the dependent variables (ψ_i) alone form a complete and independent set of initial data, the canonical conjugates of the ψ_i being expressible in terms of these data. In a suitable representation the canonical conjugates are just the complex conjugates. It then turns out that one obtains a quantum theory satisfying the Einstein-Bose statistics. This does not agree with physical experience since electrons satisfy the Fermi-Dirac statistics.

Speaking more generally, classical Poisson brackets always lead to quantum *commutation* brackets and never to *anticommutation* brackets. For this reason a description of particles satisfying the Fermi-Dirac statistics appears to be impossible with the resources of the present procedure.

However, if once the anticommutation laws in the customary form (referring to a space section of constant time) are *assumed*, our procedure enables us to express them in relativistically invariant form (referring to an arbitrary space-like section) and thus to dispense with an explicit proof of the relativistic invariance. For this purpose one need only replace the space variables x, y, z by the parameters u^1, u^2, u^3 of the space-like section and the time variable t by the time-like variable w. The same remark applies to a theory which involves Einstein-Bose particles together with Fermi-Dirac particles (Bhabha 1938; Kemmer 1938).

SUMMARY

A procedure of quantizing a dynamical continuum, developed in a previous paper, is applied to the electromagnetic field in generalized form as given by Mie and by Born. Important special cases are the Proca field of

the Heavy Electron, whose structure is similar to the linear case of Mie's theory, and the Maxwell field which is the linear case of Born's theory. In the theory of Mie, which is not gauge-invariant, the stress-energy-momentum tensor suggested by the mathematical procedure differs from the one defined in physics, but it is shown that both tensors lead to the same physical results. In the theory of Born, which is gauge-invariant, the "potential" of the field is described by a line-function $\oint \phi_i dx^i$ which has an absolute physical significance while the world vector (ϕ_i) has not. In this way the ambiguity concerning the stress-energy-momentum tensor is avoided. The well-known peculiar difficulties due to the "antisymmetry" of electrodynamics can be overcome by a clear formulation of the correct initial data in the classical theory: The present method is relativistically invariant throughout, so that an explicit proof of the invariance of the quantum conditions becomes unnecessary.

REFERENCES

Bhabha 1938 *Proc. Roy. Soc.* A, **166**, 501.
Born 1934 *Proc. Roy. Soc.* A, **143**, 410.
— 1937 *Ann. Inst. Poincaré*, **7**, 155.
Born and Infeld 1934*a* *Proc. Roy. Soc.* A, **144**, 425.
— — 1934*b* *Proc. Roy. Soc.* A, **147**, 522.
— — 1935 *Proc. Roy. Soc.* A, **150**, 141.
Dirac 1935 "The principles of quantum mechanics", 2nd ed. Oxford.
Heisenberg and Pauli 1929 *Z. Phys.* **56**, 1.
Infeld 1937 *Proc. Roy. Soc.* A, **158**, 368.
Kemmer 1938 *Proc. Roy. Soc.* A, **166**, 127.
Mie 1912*a* *Ann. Phys., Lpz.*, **37**, 511.
— 1912*b* *Ann. Phys., Lpz.*, **39**, 1.
— 1913 *Ann. Phys., Lpz.*, **40**, 1.
Pauli and Weisskopf 1934 *Helv. phys. Acta*, **7**, 709.
Proca 1936 *J. Phys. Radium* (vii), **7**, 347.
Pryce 1937 *Proc. Roy. Soc.* A, **159**, 355.
Weiss 1936 *Proc. Roy. Soc.* A, **156**, 192.
— 1938 *Proc. Roy. Soc.* A, **169**, 102.

The statistical mechanics of co-operative assemblies

By E. A. GUGGENHEIM, M.A., Sc.D.
Lecturer in Chemical Thermodynamics, Imperial College of Science and Technology, London

(*Communicated by R. H. Fowler, F.R.S.—Received* 5 *September* 1938)

INTRODUCTION

The method first used by Bethe (1935) in the statistical treatment of alloys such as β-brass, with a superlattice structure, has since been applied by other authors to assemblies of a simple type in which there is no long distance order. The object of the present paper is to extend, amend and simplify the treatment of assemblies of this simple type. It is convenient to develop the treatment in detail before considering the work of these other authors.

1. REGULAR ASSEMBLIES

The epithet *co-operative* is applied (see Fowler 1936) to an assembly when its properties cannot be expressed in terms of the properties of a large number of independent or almost independent systems (molecules, atoms, electrons, ions), because the states of any distinct system are fundamentally influenced by which states of the other systems are occupied. Considerable advances have been made recently in the statistical mechanics of co-operative assemblies, particularly of a class of co-operative assemblies with the following assumed properties:

(1) Any configuration of the assembly can be described by specifying the number and geometry of occupiable sites in the assembly and how each site is occupied.

(2) Interactions between pairs of molecules that are not nearest neighbours may be neglected and therefore the interaction energy of the systems composing the assembly can for each configuration be expressed as the sum of contributions from each pair of immediate neighbours.

(3) The internal degrees of freedom of each system are independent of its neighbours.

Possible examples of such assemblies are mixed crystals, "regular" liquid mixtures, solutions of hydrogen in palladium, monolayers adsorbed on a surface. In mixed crystals the sites are lattice-points; in monolayers on a surface the sites form a two-dimensional lattice; in solutions of hydrogen in palladium the sites are the spaces between the palladium nuclei; in liquids the sites are not fixed in space but liquid mixtures may be included in the same class as the other examples mentioned, provided the various types of molecules are sufficiently alike in size and shape that every molecule is surrounded on the average by the same number of other molecules, of whatever type. The class of liquid mixtures which satisfy these conditions corresponds roughly to the class of liquids which Hildebrand (1929) has named *regular solutions*. It therefore seems natural to apply the epithet *regular* to any assembly satisfying the criteria (1) and (2) above, and this terminology will be used here.

In two of the mentioned examples of regular assemblies, namely solutions of hydrogen in palladium and monolayers adsorbed on a crystalline surface, any site may be either empty or full. In the other three examples every site must be occupied. The distinction is, however, irrelevant for what follows.

We shall immediately proceed to develop some quite general statistical mechanical formulae for regular assemblies and postpone reference to previous work on the subject until a later stage.

2. GRAND PARTITION FUNCTIONS

The procedure which we shall adopt is to construct the grand partition function of the assembly. For a detailed description of this function and its use the reader is referred to a recent article by Fowler (1938). We denote by N the total number of sites, by N_α the number occupied *in the manner* α, by N_β the number occupied *in the manner* β, and so on. Of the quantities N_α, N_β, ... we could define at least all but one of them by saying that N_α denotes the number of sites occupied *by systems of type* α and so on. But there may be one quantity of the set N_α, N_β, ... say N_0 which denotes the number of empty sites. In this case occupied *in the manner* 0, means *empty*, but occupied in any other manner α, β, ... means occupied by systems of the type α, β, ... We further denote by $N_{\alpha\beta}$ the number of pairs of neighbouring sites the one occupied in the manner α and the other in the manner β. We denote by z the number of closest neighbours of each site. For liquids the number of closest neighbours fluctuates slightly and z must be regarded as the average value; for all the other examples z is a true constant determined by the type of lattice. We have then the necessary equalities

$$\Sigma_\alpha N_\alpha = N, \tag{2·1}$$

$$2N_{\alpha\alpha} + \sum_{\substack{\beta \\ \alpha\neq\beta}} N_{\alpha\beta} = zN_\alpha \quad \text{or} \quad \Sigma_\beta \sigma_{\alpha\beta} N_{\alpha\beta} = zN_\alpha, \tag{2·2}$$

where $\sigma_{\alpha\beta}$ is the usual symmetry number equal to 2 when α and β are identical and equal to 1 otherwise.

The grand partition function Ξ for an assembly with N sites is then defined by

$$\Xi = \Sigma_{N_{\alpha},\ldots} \Sigma_{N_{\alpha\beta},\ldots} g(N_{\alpha,\ldots}, N_{\alpha\beta,\ldots})\, e^{-\Sigma N_{\alpha\beta} w_{\alpha\beta}/kT} (\lambda_{\alpha}\phi_{\alpha})^{N_{\alpha}} (\lambda_{\beta}\phi_{\beta})^{N_{\beta}} \ldots, \qquad (2\cdot3)$$

where k denotes Boltzmann's constant and T the absolute temperature. $g(N_{\alpha,\ldots}, N_{\alpha\beta,\ldots})$ denotes the number of distinguishable configurations for the specified values of $N_{\alpha}, \ldots, N_{\alpha\beta}, \ldots$. The contribution of each $\alpha\beta$ pair of neighbours to the interaction energy is denoted by $w_{\alpha\beta}$. The partition function of a system α for all degrees of freedom and forms of energy other than the interaction energy is denoted by ϕ_{α}. We can, if we wish, include in each of $\phi_{\alpha}, \phi_{\beta}, \ldots$ an extra factor so that $w_{\alpha\alpha}, w_{\beta\beta}, \ldots$ become zero and $w_{\alpha\beta}$ becomes the excess energy of an $\alpha\beta$ pair over the mean of the energies of an $\alpha\alpha$ pair and a $\beta\beta$ pair. This is, however, merely a matter of convenience and is irrelevant to our further procedure. Finally the quantities $\lambda_{\alpha}, \lambda_{\beta}, \ldots$ are related to Gibbs' partial potentials $\mu_{\alpha}, \mu_{\beta}, \ldots$ by

$$\mu_{\alpha} = kT \log \lambda_{\alpha} \qquad (2\cdot4)$$

and the like. The appropriate thermodynamic name for λ_{α} seems to us to be the *absolute activity of* α, the ordinary *activity* a_{α} of α used by G. N. Lewis (1923) being defined by

$$\mu_{\alpha} = \mu_{\alpha}^{0}(T) + kT \log a_{\alpha}, \qquad (2\cdot5)$$

where $\mu_{\alpha}^{0}(T)$ is arbitrarily fixed at each temperature.

The properties of the grand partition function which we shall require to use are the following. For the given values of $T, \lambda_{\alpha}, \lambda_{\beta}, \ldots$ the probability of some configuration with specified values of $N_{\alpha}, N_{\beta}, \ldots$ and $N_{\alpha\beta}, \ldots$ is proportional to $g(N_{\alpha,\ldots}, N_{\alpha\beta,\ldots}) e^{-\Sigma N_{\alpha\beta} w_{\alpha\beta}/kT} (\lambda_{\alpha}\phi_{\alpha})^{N_{\alpha}} (\lambda_{\beta}\phi_{\beta})^{N_{\beta}} \ldots$. This enables one to write down all average properties. In particular one finds that the average (equilibrium) value of N_{α} for given $T, \lambda_{\alpha}, \ldots$ is given by

$$\overline{N}_{\alpha} = \lambda_{\alpha} \frac{\partial \log \Xi}{\partial \lambda_{\alpha}}, \qquad (2\cdot6)$$

and the average (equilibrium) value of $N_{\alpha\beta}$ is given by

$$\overline{N}_{\alpha\beta} = -kT \frac{\partial \log \Xi}{\partial w_{\alpha\beta}}. \qquad (2\cdot7)$$

Strictly $\phi_{\alpha}, \phi_{\beta}, \ldots$ depend on the volume V, as well as the temperature, and $w_{\alpha\beta}, \ldots$ depend on the volume V of the assembly. (For volume, read area, in the case of a monolayer assembly). The pressure P (two dimensional in the case of the monolayer) is then given by

$$P = kT\frac{\partial \log \Xi}{\partial V} = \frac{kT}{V}\log \Xi. \tag{2·8}$$

It is convenient to re-write this in the form

$$P = \frac{NkT}{V}\log \Xi^{1/N} = \frac{kT}{\mathrm{v}}\log \Xi^{1/N}, \tag{2·9}$$

where $\mathrm{v} = V/N$ denotes the volume per system and $\Xi^{1/N}$ is for all relevant large values of N independent of N, that is to say independent of the size of the assembly. It is usually more profitable to regard (2·9) as a condition of internal equilibrium *at a given* P rather than as a relation determining P. In particular P is usually negligibly small compared with kT/v and (2·9) then reduces to

$$\Xi^{1/N} \sim 1. \tag{2·10}$$

This relation will play an important part when we come to contrast our formulae with those used previously by various other authors.

For brevity's sake it is convenient to introduce the quantities $\xi_\alpha, \xi_\beta, \ldots, \eta_{\alpha\beta}, \ldots$ defined by

$$\xi_\alpha = \lambda_\alpha \phi_\alpha, \tag{2·11}$$

$$\eta_{\alpha\beta} = e^{-w_{\alpha\beta}/kT}. \tag{2·12}$$

The formula for Ξ now becomes

$$\Xi = \Sigma_{N_\alpha,\ldots}\Sigma_{N_{\alpha\beta},\ldots}\, g(N_\alpha,\ldots,N_{\alpha\beta},\ldots)\,\eta_{\alpha\beta}^{N_{\alpha\beta}}\ldots\xi_\alpha^{N_\alpha}\xi_\beta^{N_\beta}\ldots, \tag{2·13}$$

and the relations (2·6), (2·7) become

$$\bar{N}_\alpha = \xi_\alpha\frac{\partial \log \Xi}{\partial \xi_\alpha}, \tag{2·14}$$

$$\bar{N}_{\alpha\beta} = \eta_{\alpha\beta}\frac{\partial \log \Xi}{\partial \eta_{\alpha\beta}}. \tag{2·15}$$

In deriving averages the error is negligible if one replaces the series for Ξ by its greatest term. If this term is Ξ^* defined by

$$\Xi^* = g(N_\alpha^*,\ldots,N_{\alpha\beta}^*,\ldots)\,\eta_{\alpha\beta}^{N_{\alpha\beta}^*}\ldots\xi_\alpha^{N_\alpha^*}\xi_\beta^{N_\beta^*}\ldots, \tag{2·16}$$

then the average values of N_α, N_β, ... and $N_{\alpha\beta}$, ... are given by

$$\bar{N}_\alpha = N_\alpha^*, \tag{2·17}$$

$$\bar{N}_{\alpha\beta} = N_{\alpha\beta}^*. \tag{2·18}$$

If then one could determine N_α^*, ... and $N_{\alpha\beta}^*$, ... one could construct Ξ^* and could so determine all the thermodynamic properties.

3. FORMAL CONSTRUCTIONS OF Ξ

To make use of formula (2·13) for Ξ we should require to evaluate the coefficients $g(N_{\alpha}, \ldots, N_{\alpha\beta}, \ldots)$ while to make use of (2·16) we should require to evaluate $N^*_{\alpha}, \ldots$ and $N^*_{\alpha\beta}, \ldots$. As there is no straightforward way of doing either, it is necessary to find some special device for evaluating Ξ.

We define the quantities ξ, η, γ by

$$\xi^N = \Pi_{\alpha} \xi_{\alpha}^{N^*_{\alpha}}, \tag{3·1}$$

$$\eta^{\frac{1}{2}zN} = \Pi_{\alpha\beta} \eta_{\alpha\beta}^{N^*_{\alpha\beta}}, \tag{3·2}$$

$$\gamma^N = g(N^*_{\alpha}, \ldots, N^*_{\alpha\beta}, \ldots). \tag{3·3}$$

Thus ξ is the geometrical mean of the factor of the type $\xi_{\alpha}, \xi_{\beta}, \ldots$ contributed by each site to Ξ^*; η is the geometrical mean of the factor of the type $\eta_{\alpha\beta}, \ldots$ contributed by each pair of neighbouring sites to Ξ^*; γ is the geometrical mean of the factor contributed by each site to $g(N^*_{\alpha}, \ldots, N^*_{\alpha\beta}, \ldots)$. We also define η_{α} as the geometrical mean of the factor of the type $\eta_{\alpha\beta}$ contributed to Ξ^* by each pair of neighbouring sites one of which is specified as being occupied in the manner α.

Using these definitions we can express Ξ^* or Ξ in a number of alternative forms amongst which are

$$\Xi = \Xi^* = \gamma^N \xi^N \eta^{\frac{1}{2}zN}, \tag{3·4}$$

$$\Xi = \Xi_1 = \Sigma_{\alpha} \xi_{\alpha} \eta_{\alpha}^{z} \gamma^{N-1} \xi^{N-1} \eta^{\frac{1}{2}zN-z}, \tag{3·5}$$

$$\Xi = \Xi_2 = \Sigma_{\alpha} \Sigma_{\beta} \xi_{\alpha} \eta_{\alpha}^{z-1} \xi_{\beta} \eta_{\beta}^{z-1} \eta_{\alpha\beta} \gamma^{N-2} \xi^{N-2} \eta^{\frac{1}{2}zN-2z+1}, \tag{3·6}$$

$$\Xi = \Xi_{z+1} = \Sigma_{\alpha} \xi_{\alpha} (\Sigma_{\beta} \xi_{\beta} \eta_{\beta}^{z-1} \eta_{\alpha\beta})^z \gamma^{N-z-1} \xi^{N-z-1} \eta^{\frac{1}{2}zN-z^2}. \tag{3·7}$$

The expressions Ξ^*, Ξ_1, Ξ_2, Ξ_{z+1} have by definition equal values, each being equal to $e^{PV/kT}$, but differences in form corresponding to differences in the physical points of view. The form of Ξ_1 is such that the assembly is regarded as one site, which we may call the central site, and a residue which forms an external field as far as the central site is concerned. The form of Ξ_2 is such that the assembly is regarded as a single pair of neighbouring sites, and a residue which forms an external field as far as the single pair of neighbouring sites is concerned. The form of Ξ_{z+1} is such that the assembly is regarded as a group of a central site and its z neighbours, that is $z+1$ sites in all, and a residue.

We assume that all the formulae of the present section can be made mutually consistent†, and valuable conclusions can be drawn from this consistency when we apply the formulae of § 2. For example, we can derive alternative formulae for θ_α, the fraction of sites occupied in the manner α or the probability that a given site is occupied in the manner α, by applying (2·14) either to Ξ_1 regarded as the Ξ for a single site in a field due to the remaining sites or to Ξ_2 regarded as the Ξ for a pair of neighbouring sites in a field due to the remaining sites. We thus obtain on the one hand using (3·5)

---

† The nature of this assumption is further considered in § 10.

$$\theta_\alpha = \frac{\overline{N}_\alpha}{N} = \xi_\alpha \frac{\partial \log \Xi_1}{\partial \xi_\alpha} = \frac{\xi_\alpha \eta_\alpha^z}{\gamma \xi \eta^z}, \tag{3·8}$$

and on the other using (3·6)

$$\theta_\alpha = \frac{\overline{N}_\alpha}{N} = \tfrac{1}{2}\xi_\alpha \frac{\partial \log \Xi_2}{\partial \xi_\alpha} = \frac{\xi_\alpha \eta_\alpha^{z-1} \Sigma_\tau \xi_\tau \eta_\tau^{z-1} \eta_{\alpha\tau}}{\gamma^2 \xi^2 \eta^{2z-1}}. \tag{3·9}$$

In order that (3·8) and (3·9) may be mutually consistent, we must have

$$\eta_\alpha = \frac{\Sigma_\tau \xi_\tau \eta_\tau^{z-1} \eta_{\alpha\tau}}{\gamma \xi \eta^{z-1}}, \tag{3·10}$$

and this is also sufficient to ensure the consistency of (3·7) with (3·5) and (3·6). For the mutual consistency of (3·4) and (3·5) it is further necessary that

$$\Sigma_\tau \xi_\tau \eta_\tau^z = \gamma \xi \eta^z. \tag{3·11}$$

We might alternatively evaluate θ_α from the relation

$$\theta_\alpha = \frac{\overline{N}_\alpha}{N} = \frac{1}{z+1} \xi_\alpha \frac{\partial \log \Xi_{z+1}}{\partial \xi_\alpha}, \tag{3·12}$$

using formula (3·7) for Ξ_{z+1}. If we do this and make use of (3·10), we obtain after simplification just (3·8) and so obtain no new relations analogous to (3·10). Nor do we obtain anything new by considering groups of sites intermediate between a pair and the group of $z+1$. More complicated groups leading to more accurate approximations are considered in § 10.

From (3·8) we deduce using (3·11)

$$\frac{\theta_\alpha}{\xi_\alpha \eta_\alpha^z} = \frac{\theta_\beta}{\xi_\beta \eta_\beta^z} = \ldots = \frac{1}{\Sigma_\tau \xi_\tau \eta_\tau^z} = \frac{1}{\gamma \xi \eta^z}, \tag{3·13}$$

and from (3·10) we deduce

$$\frac{\eta_\alpha}{\Sigma_\tau \xi_\tau \eta_\tau^{z-1} \eta_{\alpha\tau}} = \frac{\eta_\beta}{\Sigma_\tau \xi_\tau \eta_\tau^{z-1} \eta_{\beta\tau}} = \ldots. \tag{3·14}$$

In (3·14) we have a set of simultaneous equations in number sufficient to determine the ratios $\eta_\alpha : \eta_\beta : \ldots$ as functions of the known quantities $\xi_\alpha, \ldots$ and $\eta_{\alpha\beta}, \ldots$ and so by (3·13) to determine $\theta_\alpha, \theta_\beta, \ldots$ as functions of $\xi_\alpha, \ldots$

and $\eta_{\alpha\beta}$, ... Since however the equations (3·14) are of degree z in $\eta_\alpha, \eta_\beta, \ldots$, they are intractable and this method of attack cannot generally lead to explicit formulae for $\theta_\alpha, \theta_\beta, \ldots$.

4. Quasi-chemical equilibria

Just as we have obtained formulae for θ_α by using (2·14), so by using (2·15) we can obtain formulae for $\theta_{\alpha\beta}$; the fraction of pairs of neighbouring sites occupied the one in the manner α, the other in the manner β. It is simplest to combine (2·15) with (3·6) for Ξ_2 regarded as the Ξ for a pair of sites in the field due to the remaining sites. We thus obtain

$$\theta_{\alpha\beta} = \frac{\bar{N}_{\alpha\beta}}{\frac{1}{2}zN} = \eta_{\alpha\beta}\frac{\partial \log \Xi_2}{\partial \eta_{\alpha\beta}} = \frac{2}{\sigma_{\alpha\beta}}\frac{\xi_\alpha \eta_\alpha^{z-1} \xi_\beta \eta_\beta^{z-1} \eta_{\alpha\beta}}{\gamma^2 \xi^2 \eta^{2z-1}}, \tag{4·1}$$

where $\sigma_{\alpha\beta}$ is a symmetry number equal to 2 when α and β are identical and 1 otherwise. From (4·1) we can deduce a number of relations such as

$$\frac{\theta_{\alpha\beta}\theta_{\gamma\delta}}{\theta_{\alpha\gamma}\theta_{\beta\delta}} = \frac{\eta_{\alpha\beta}\eta_{\gamma\delta}}{\eta_{\alpha\gamma}\eta_{\beta\delta}} \Big/ \frac{\sigma_{\alpha\beta}\sigma_{\gamma\delta}}{\sigma_{\alpha\gamma}\sigma_{\beta\delta}}, \tag{4·2}$$

and in particular

$$\frac{\theta_{\alpha\beta}^2}{\theta_{\alpha\alpha}\theta_{\beta\beta}} = \frac{4\eta_{\alpha\beta}^2}{\eta_{\alpha\alpha}\eta_{\beta\beta}} = 4e^{-(2w_{\alpha\beta}-w_{\alpha\alpha}-w_{\beta\beta})/kT}. \tag{4·3}$$

It is remarkable that these relations, from which the quantities $\eta_\alpha, \eta_\beta, \ldots$ as well as ξ, η, γ have been completely eliminated, have been obtained from (3·6) without any use being made of (3·10). It will be noticed that these relations have the simple form of a quasi-chemical equilibrium. We shall now use them to treat the problem by an entirely different method.

5. Free energy and total energy

The ordinary partition function, corresponding to the grand partition function (2·3), for an assembly with given $N_\alpha, N_\beta, \ldots$, instead of given $\lambda_\alpha, \lambda_\beta, \ldots$ as previously, is

$$\phi_\alpha^{N_\alpha}\phi_\beta^{N_\beta}\ldots\Sigma_{N_{\alpha\beta},\ldots}\, g(N_{\alpha\beta,\ldots})\, e^{-\Sigma N_{\alpha\beta}w_{\alpha\beta}/kT}, \tag{5·1}$$

where $g(N_{\alpha\beta,\ldots})$ denotes the number of distinguishable configurations with the specified values of $N_{\alpha\beta,\ \ldots}$ which are subject to the necessary relations (2·2). We now define quantities $\bar{\bar{N}}_{\alpha\beta,\ \ldots}$ such that whatever values be assigned to $w_{\alpha\beta,\ldots}$

$$e^{-\Sigma\bar{N}_{\alpha\beta}w_{\alpha\beta}/kT}\,\Sigma_{N_{\alpha\beta},\ldots}\, g(N_{\alpha\beta,\ldots}) = \Sigma_{N_{\alpha\beta},\ldots}\, g(N_{\alpha\beta,\ldots})\, e^{-\Sigma N_{\alpha\beta}w_{\alpha\beta}/kT}, \tag{5·2}$$

while the true average values $\bar{N}_{\alpha\beta,\ldots}$ are given by

$$\bar{N}_{\alpha\beta}\Sigma_{N_{\alpha\beta},\ldots}g(N_{\alpha\beta},\ldots)\,e^{-\Sigma N_{\alpha\beta}w_{\alpha\beta}/kT} = \Sigma_{N_{\alpha\beta},\ldots}N_{\alpha\beta}\,g(N_{\alpha\beta},\ldots)\,e^{-\Sigma N_{\alpha\beta}w_{\alpha\beta}/kT}. \quad (5\cdot3)$$

By taking logarithms of (5·2) and differentiating with respect to T, we can verify that

$$\bar{N}_{\alpha\beta} = \frac{\partial(\bar{\bar{N}}_{\alpha\beta}/T)}{\partial(1/T)}. \quad (5\cdot4)$$

The free energy F of the assembly according to (5·1) is given by

$$F = -\Sigma_\alpha N_\alpha kT\log\phi_\alpha - kT\log\Sigma_{N_{\alpha\beta},\ldots}g(N_{\alpha\beta},\ldots)\,e^{-\Sigma N_{\alpha\beta}w_{\alpha\beta}/kT}. \quad (5\cdot5)$$

Using (5·2), we can re-write this as

$$\begin{aligned} F &= -\Sigma_\alpha N_\alpha kT\log\phi_\alpha - kT\log\Sigma_{N_{\alpha\beta},\ldots}g(N_{\alpha\beta},\ldots)+\Sigma_{\alpha\beta}\bar{\bar{N}}_{\alpha\beta}w_{\alpha\beta} \\ &= -\Sigma_\alpha N_\alpha kT\log\phi_\alpha - kT\log(N!/N_\alpha!\,N_\beta!\ldots)+\Sigma_{\alpha\beta}\bar{\bar{N}}_{\alpha\beta}w_{\alpha\beta} \\ &= \Sigma_\alpha N_\alpha kT\log\frac{N_\alpha}{N\phi_\alpha}+\Sigma_{\alpha\beta}\bar{\bar{N}}_{\alpha\beta}w_{\alpha\beta}. \end{aligned} \quad (5\cdot6)$$

We see that the contribution of the interaction energies to the free energy of the assembly is

$$F^{\text{int}} = \Sigma_{\alpha\beta}\bar{\bar{N}}_{\alpha\beta}w_{\alpha\beta}. \quad (5\cdot7)$$

The corresponding contribution to the total energy E is

$$E^{\text{int}} = \frac{\partial(F^{\text{int}}/T)}{\partial(1/T)} = \Sigma_{\alpha\beta}\bar{N}_{\alpha\beta}w_{\alpha\beta}, \quad (5\cdot8)$$

according to (5·4).

If then we can determine $\bar{\bar{N}}_{\alpha\beta},\ \ldots$, or alternatively $\bar{N}_{\alpha\beta},\ \ldots$ since the two sets of quantities are related by (5·4), as functions of T, N_α, N_β, …, we shall have an explicit formula (5·6) for the free energy as a function of T, N_α, N_β, … and from this we can derive all the thermodynamic functions of the assembly. But according to (4·3) we have

$$\frac{\bar{\bar{N}}_{\alpha\beta}^2}{\bar{\bar{N}}_{\alpha\alpha}\bar{\bar{N}}_{\beta\beta}} = 4e^{-(2w_{\alpha\beta}-w_{\alpha\alpha}-w_{\beta\beta})/kT}, \quad (5\cdot9)$$

and according to (2·2)

$$2\bar{\bar{N}}_{\alpha\alpha}+\sum_{\substack{\beta \\ \alpha\neq\beta}}\bar{\bar{N}}_{\alpha\beta} = zN_\alpha. \quad (5\cdot10)$$

There are as many equations of the type (5·9) as quantities $\bar{\bar{N}}_{\alpha\beta}$ $(\alpha\neq\beta)$ and as many equations of the type (5·10) as quantities $\bar{\bar{N}}_{\alpha\alpha}$. Hence there are just sufficient equations (5·9) and (5·10) together to determine all the $\bar{\bar{N}}_{\alpha\alpha}$ and $\bar{\bar{N}}_{\alpha\beta}$ as functions of N_α, N_β, ….

The number of simultaneous equations to be solved is greater than in the method of § 3, but the equations are all quadratic or linear, whereas those in § 3 were of order z. In practice the equations of the present section are by far the more tractable.

6. COMPARISON WITH PERFECT ASSEMBLIES

We define a *perfect* assembly as one such that

$$2w_{\alpha\beta} = w_{\alpha\alpha} + w_{\beta\beta} \quad (\text{all } \alpha, \beta). \tag{6·1}$$

It is usual to describe a regular assembly by comparing its properties with those of a perfect assembly chosen as standard, but in making this comparison the perfect assembly used as standard is chosen in a different way according as sites may or may not remain empty.

Let us first consider assemblies such as mixed crystals and regular solutions in which all the sites must be occupied. We shall call these regular assemblies of class 1. The perfect assembly chosen as standard is then one in which $w_{\alpha\alpha}$, $w_{\beta\beta}$, ... have the same values as in the actual assembly, while the $w_{\alpha\beta}$, ... are determined by (6·1). If then we use the symbol Δ to denote the excess of the value for the actual assembly over its value for the chosen standard perfect assembly, we have

$$\Delta F = \Sigma_{\alpha\beta} \overline{N}_{\alpha\beta}(w_{\alpha\beta} - \tfrac{1}{2}w_{\alpha\alpha} - \tfrac{1}{2}w_{\beta\beta}), \tag{6·2}$$

$$\Delta E = \Sigma_{\alpha\beta} \overline{N}_{\alpha\beta}(w_{\alpha\beta} - \tfrac{1}{2}w_{\alpha\alpha} - \tfrac{1}{2}w_{\beta\beta}). \tag{6·3}$$

It is usual and convenient to include in the partition functions ϕ_α, ϕ_β, ... factors so that $w_{\alpha\alpha}$, $w_{\beta\beta}$, ... become zero. Formulae (6·2), (6·3) then reduce to

$$\Delta F = \Sigma_{\alpha\beta} \overline{N}_{\alpha\beta} w_{\alpha\beta}, \quad (w_{\alpha\alpha} = \ldots = 0), \tag{6·4}$$

$$\Delta E = \Sigma_{\alpha\beta} \overline{N}_{\alpha\beta} w_{\alpha\beta}, \quad (w_{\alpha\alpha} = \ldots = 0). \tag{6·5}$$

According to this convention the quantities $\eta_{\alpha\alpha}$, $\eta_{\beta\beta}$, ... become equal to unity.

In assemblies such as solutions of hydrogen in palladium and monolayers on a crystal surface some of the sites may be empty. We shall call these regular assemblies of class 2. For these assemblies we use the subscript 0 to denote an empty site. The perfect assembly chosen as standard is now one in which $w_{\alpha 0}$, $w_{\beta 0}$, ... have the same values as in the actual assembly, the $w_{\alpha\alpha}$ being determined by

$$w_{\alpha\alpha} + w_{00} = w_{\alpha\alpha} = 2w_{\alpha 0}, \tag{6·6}$$

since w_{00} is obviously zero, and the remaining $w_{\alpha\beta}$ being given by

$$w_{\alpha\beta} = \tfrac{1}{2}(w_{\alpha\alpha}+w_{\beta\beta}) = w_{\alpha0}+w_{\beta0}. \tag{6·7}$$

When we now compare the properties of the actual assembly with those of the chosen standard perfect assembly, we find

$$\Delta F = \Sigma_{\alpha\beta}\overline{N}_{\alpha\beta}(w_{\alpha\beta}-w_{\alpha0}-w_{\beta0}), \tag{6·8}$$

$$\Delta E = \Sigma_{\alpha\beta}\overline{N}_{\alpha\beta}(w_{\alpha\beta}-w_{\alpha0}-w_{\beta0}). \tag{6·9}$$

It is now convenient to include in the partition functions $\phi_\alpha, \phi_\beta, \ldots$ factors so that $w_{\alpha0}, w_{\beta0}, \ldots$ become zero. Formulae (6·8) and (6·9) then reduce to

$$\Delta F = \Sigma_{\alpha\beta}\overline{N}_{\alpha\beta}w_{\alpha\beta}, \quad (w_{\alpha0}=\ldots=0), \tag{6·10}$$

$$\Delta E = \Sigma_{\alpha\beta}\overline{N}_{\alpha\beta}w_{\alpha\beta}, \quad (w_{\alpha0}=\ldots=0). \tag{6·11}$$

These are of the same form as (6·4) and (6·5) but it is now $\eta_{\alpha0}, \eta_{\beta0}, \ldots$ that become unity.

The essential similarity of the two classes of regular assemblies is apt to be obscured by the two different conventions, one leading to $\eta_{\alpha\alpha}, \eta_{\beta\beta}, \ldots$ equal to unity, the other leading to $\eta_{\alpha0}, \eta_{\beta0}, \ldots$ equal to unity. We have therefore avoided using either convention in the earlier stages.

7. SITES OCCUPIABLE IN TWO MANNERS

All previous work on this subject has been confined to assemblies whose sites can be occupied in only two manners. We therefore re-write some of our formulae for this special case in order to compare them with formulae previously used. We now distinguish between the two types of regular assemblies.

For assemblies of class 1 such that each site must be occupied by either a system α or a system β, we define ϕ_α, ϕ_β so that $w_{\alpha\alpha} = w_{\beta\beta} = 0$ or $\eta_{\alpha\alpha} = \eta_{\beta\beta} = 1$. Formulae (3·8) and (3·11) now reduce to

$$\theta_\alpha = \frac{\xi_\alpha\eta_\alpha^z}{\gamma\xi\eta^z} = \frac{\xi_\alpha\eta_\alpha^z}{\xi_\alpha\eta_\alpha^z+\xi_\beta\eta_\beta^z}, \quad \theta_\beta = \frac{\xi_\beta\eta_\beta^z}{\gamma\xi\eta^z} = \frac{\xi_\beta\eta_\beta^z}{\xi_\alpha\eta_\alpha^z+\xi_\beta\eta_\beta^z}, \tag{7·1}$$

while formula (3·10) reduces to

$$\eta_\alpha = \frac{\xi_\alpha\eta_\alpha^{z-1}+\xi_\beta\eta_\beta^{z-1}\eta_{\alpha\beta}}{\gamma\xi\eta^{z-1}}, \quad \eta_\beta = \frac{\xi_\alpha\eta_\alpha^{z-1}\eta_{\alpha\beta}+\xi_\beta\eta_\beta^{z-1}}{\gamma\xi\eta^{z-1}}. \tag{7·2}$$

For assemblies of class 2 such that each site may be full or empty, we use the subscript 0 for the empty state and the subscript α for the full. We set

$\lambda_0 = 1$, $\phi_0 = 1$, $\xi_0 = 1$, $\eta_{00} = 1$ in order that the free energy of a completely empty assembly may be zero. We include in ϕ_α a factor so that $w_{\alpha 0} = 0$ or $\eta_{\alpha 0} = 1$. Since η_0 is an averaged value of η_{00} and $\eta_{\alpha 0}$ which are both unity, it follows that $\eta_0 = 1$ also. Formulae (3·8) and (3·10) now reduce to

$$\theta_\alpha = \frac{\xi_\alpha \eta_\alpha^z}{\gamma \xi \eta^z} = \frac{\xi_\alpha \eta_\alpha^z}{\xi_\alpha \eta_\alpha^z + 1}, \quad \theta_0 = 1 - \theta_\alpha = \frac{1}{\xi_\alpha \eta_\alpha^z + 1}, \tag{7·3}$$

$$\eta_\alpha = \frac{\xi_\alpha \eta_\alpha^{z-1} \eta_{\alpha\alpha} + 1}{\gamma \xi \eta^{z-1}}, \quad \eta_0 = \frac{\xi_\alpha \eta_\alpha^{z-1} + 1}{\gamma \xi \eta^{z-1}} = 1. \tag{7·4}$$

8. Previous treatment

The methods of § 5 have been used previously by the author (1935) in an attempt to determine the equilibrium properties of regular solutions. Formulae of the quasi-chemical equilibrium type like (5·9) were there assumed intuitively. They have since been derived for regular solutions by Rushbrooke (1938) who used the method due to Bethe (1935) which will be discussed in the next section. At the same time Rushbrooke has pointed out an error in the author's paper in that he applied the equations of the quasi-chemical equilibrium type to the $\overline{\overline{N}}_{\alpha\beta}$ instead of to the $\overline{N}_{\alpha\beta}$. The formulae of the author's previous paper are therefore wrong, but when the error is corrected the method of attack, which is that used in § 5, is probably the most convenient for determining the equilibrium properties of any regular assembly.

9. Bethe's method

For convenience we have postponed to this stage any detailed discussion of Bethe's work. The method used by Bethe (1935) in his treatment of metallic alloys has since been applied to monolayers by Peierls (1936) and by Wang (1937), to solutions of hydrogen in palladium by Lacher (1937) and to regular solutions by Rushbrooke (1938). The reader will have noticed the similarity of the method of the treatment described here to that of Bethe, out of which it has in fact been developed. There is no need to dwell on this similarity and we shall confine our discussion to differences.

As already mentioned, all previous work has been confined to cases where the sites can be occupied in only two ways. This restriction is removed in the present treatment. We have also avoided any reference to "right" and "wrong" manners of occupation of a site, these conceptions being irrelevant to assemblies in which there is no long range order. We have also deliberately avoided any consideration of the sign of $w_{\alpha\beta} - \frac{1}{2}(w_{\alpha\alpha} + w_{\beta\beta})$, since this is irrelevant to the development. It is true that when this is positive, but not when it is negative, we shall find that at sufficiently low temperatures $\partial\xi_\alpha/\partial N_\alpha$ or $\partial\mu_\alpha/\partial N_\alpha$ can become negative for certain values of N_α/N_β. Such phases will be unstable and will split into two phases and we have critical phenomena. But this follows naturally from the formula for F or μ_α or ξ_α as functions of N_α/N_β and there is no point in giving separate developments of the formulae for the two cases $w_{\alpha\beta} > \frac{1}{2}(w_{\alpha\alpha}+w_{\beta\beta})$ and $w_{\alpha\beta} < \frac{1}{2}(w_{\alpha\alpha}+w_{\beta\beta})$.

As far as the above comments are concerned, the formulae of Bethe, Peierls, Wang, Lacher and Rushbrooke might be special examples of our formulae, but this is not the case. There is a real difference, though, as we shall see, not a serious one. If for assemblies of class 1 we define a quantity ϵ' by

$$\epsilon' = (\eta_\beta/\eta_\alpha)^{z-1}, \tag{9·1}$$

we can derive from (7·1) and (7·2)

$$\frac{\theta_\beta}{\theta_\alpha}\frac{\xi_\alpha}{\xi_\beta} = \epsilon'^{z/(z-1)}, \tag{9·2}$$

$$\epsilon' = \left(\frac{\xi_\alpha \eta_{\alpha\beta} + \epsilon'\xi_\beta}{\xi_\alpha + \epsilon'\eta_{\alpha\beta}\xi_\beta}\right)^{z-1}. \tag{9·3}$$

Formulae (9·2) and (9·3) are identical with the formulae for regular solutions obtained by Rushbrooke using Bethe's method. Similarly, if for assemblies of class 2 we define a quantity ζ by

$$\zeta = \eta_\alpha^{z-1}, \tag{9·4}$$

we can derive from (7·3) and (7·4)

$$\frac{\theta_\alpha}{1-\theta_\alpha} = \xi_\alpha \zeta^{z/(z-1)}, \tag{9·5}$$

$$\zeta = \left(\frac{1+\zeta\xi_\alpha\eta_{\alpha\alpha}}{1+\zeta\xi_\alpha}\right)^{z-1}. \tag{9·6}$$

If we omitted the subscripts α, (9·5), (9·6) would become identical with the formulae for monolayers obtained by Peierls using Bethe's method. We see then that we obtain the same formulae for $\theta_\alpha/\theta_\beta$ or for $\theta_\alpha/(1-\theta_\alpha)$ as are obtained by Bethe's method. One difference between the two methods is that whereas Bethe's method deals throughout with a group of $z+1$ sites we have obtained all the essential formulae by considering only a pair of sites; we have included certain formulae referring to a group of $z+1$ sites merely to show the absence of any inconsistency. The other essential difference between the two methods is that we have defined η_α and η_β in such a way that their values are unambiguous, whereas Bethe's method involves only the ratio η_β/η_α. It is stated explicitly by Rushbrooke, and it is tacitly implied in the work of Bethe and Peierls, that only this ratio has any significance. It is true that the absolute values of η_α, η_β are not required to determine θ_α, θ_β; but these values do matter if Ξ is to be the grand partition function with a value equal to $e^{NP\mathrm{v}/kT}$. If for simplicity we confine ourselves to the usual case when $P\mathrm{v} \ll kT$, then we have the approximate equality (2·10). According to (3·4) and (3·5) this requires

$$\Sigma_\alpha \xi_\alpha \eta_\alpha^z = \gamma\xi\eta^z \sim 1, \tag{9·7}$$

or in the simple case of only two types of systems

$$\xi_\alpha \eta_\alpha^z + \xi_\beta \eta_\beta^z \sim 1. \tag{9·8}$$

The relation (9·8) determines the absolute values of η_α, η_β when their ratio is determined by (7·2).

10. NATURE OF APPROXIMATION

The formulae, which have been derived, are accurate only in so far as there is no long distance order. But, if the absence of long distance order means that the probability for a given site to be occupied in a particular way is independent of the manner of occupation of all other sites other than its neighbours, then there can strictly never be complete absence of long distance order. For suppose a certain site is specified as occupied in a certain way, then the probabilities of occupation in various ways of its immediate neighbours are immediately affected; but this affects the probabilities of the next neighbours and these in turn those of their neighbours and so on. Thus strictly the probabilities of occupation of each site are linked at least remotely to those of every other site however distant.

The analytical form of the approximation consists in using the same factors η_α, η_β, ... in whatever manner the other sites, other than immediate neighbours, be occupied. We have seen that subject to this approximation we can use the same value for these factors whether we consider a single site, a pair of sites or a group of a central site with its z neighbours. We should continue to obtain mutually consistent equations for even larger groups of sites provided these contain only branches but no rings (in the terminology of organic chemisty). But as soon as we consider a group containing a closed ring we obtain different results.

For simplicity let us consider a square array with two manners of occupation α and β. Then we have for a single site

$$\Xi_1 = (\xi_\alpha \eta_\alpha^4 + \xi_\beta \eta_\beta^4)\, \gamma^{N-1} \xi^{N-1} \eta^{2N-4}. \tag{10·1}$$

For a pair of sites we have

$$\Xi_2 = \{\xi_\alpha \eta_\alpha^3(\xi_\alpha \eta_\alpha^3 \eta_{\alpha\alpha} + \xi_\beta \eta_\beta^3 \eta_{\alpha\beta}) + \xi_\beta \eta_\beta^3(\xi_\alpha \eta_\alpha^3 \eta_{\alpha\beta} + \xi_\beta \eta_\beta^3 \eta_{\beta\beta})\}\, \gamma^{N-2} \xi^{N-2} \eta^{2N-7}. \tag{10·2}$$

For a central site with its four neighbours we have

$$\Xi_5 = \{\xi_\alpha(\xi_\alpha \eta_\alpha^3 \eta_{\alpha\alpha} + \xi_\beta \eta_\beta^3 \eta_{\alpha\beta})^4 + \xi_\beta(\xi_\alpha \eta_\alpha^3 \eta_{\alpha\beta} + \xi_\beta \eta_\beta^3 \eta_{\beta\beta})^4\}\, \gamma^{N-5} \xi^{N-5} \eta^{2N-16}. \tag{10·3}$$

As we saw in § 3 these forms for Ξ are all mutually consistent with the same values for η_α, η_β. If however we construct Ξ_4 for a group of four sites forming a square, we find

$$\begin{aligned}\Xi_4 = \{&(\xi_\alpha \eta_\alpha^2)^4 \eta_{\alpha\alpha}^4 + 4(\xi_\alpha \eta_\alpha^2)^3 \xi_\beta \eta_\beta^2 \eta_{\alpha\alpha}^2 \eta_{\alpha\beta}^2 + (\xi_\alpha \eta_\alpha^2)^2 (\xi_\beta \eta_\beta^2)^2 (4\eta_{\alpha\alpha} \eta_{\alpha\beta}^2 \eta_{\beta\beta} + 2\eta_{\alpha\beta}^4) \\ &+ 4\xi_\alpha \eta_\alpha^2 (\xi_\beta \eta_\beta^2)^3 \eta_{\alpha\beta}^2 \eta_{\beta\beta}^2 + (\xi_\beta \eta_\beta^2)^4 \eta_{\beta\beta}^4\}\, \gamma^{N-4} \xi^{N-4} \eta^{2N-12}.\end{aligned} \tag{10·4}$$

This Ξ_4 is not equal to Ξ_1, Ξ_2, Ξ_5 with the same values of η_α, η_β. This inconsistency exposes the nature of the approximation.

It is almost obvious that the more closely the sites are packed, the smaller will be the least group which manifests the inconsistency and the more serious the approximation of our treatment. For a closest packed lattice our constructions of Ξ_1 and Ξ_2 are consistent with each other but not with that for Ξ_{z+1}. Fortunately Rushbrooke's work on regular solutions shows that for the simple types of assembly under consideration the error due to our approximation is not serious even for a closest packed lattice.

The author is grateful to Professor C. G. Darwin, F.R.S. and Professor R. H. Fowler, F.R.S., for their interest and advice and to Mr G. S. Rushbrooke for helpful discussions and constructive criticism.

SUMMARY

A regular assembly is defined as a co-operative assembly with certain simple properties analogous to those of regular solutions. The definition covers an important class of assemblies including, in addition to regular solutions, mixed crystals, solutions of hydrogen in palladium and monolayers on a surface. A general statistical treatment of regular assemblies is given, the approximation involved being equivalent to that of Bethe. The

treatment given may be considered as generalizing, amending and to some extent simplifying the treatment of various examples of regular assemblies due to Peierls, Wang, Lacher and Rushbrooke, using Bethe's method.

REFERENCES

Bethe, H. 1935 *Proc. Roy. Soc.* A, **150**, 552–75.
Fowler, R. H. 1936 "Statistical Mechanics", 2nd ed., § 21·1.
— 1938 *Proc. Camb. Phil. Soc.* **34**, 382–391.
Guggenheim, E. A. 1935 *Proc. Roy. Soc.* A, **148**, 304–12.
Hildebrand, J. 1929 *J. Amer. Chem. Soc.* **51**, 66–80.
Lacher, J. R. 1937 *Proc. Camb. Phil. Soc.* **33**, 518–523.
Lewis, G. N. and Randall, M. 1923 "Thermodynamics", p. 255.
Peierls, R. 1936 *Proc. Camb. Phil. Soc.* **32**, 471–76.
Rushbrooke, G. S. 1938 *Proc. Roy. Soc.* A, **166**, 296–315.
Wang, J. 1937 *Proc. Roy. Soc.* A, **161**, 127–40.

The Bakerian Lecture

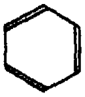
The structure of benzene

By C. K. Ingold, F.R.S.

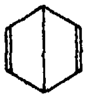
(*Read* 16 *June* 1938.—*Received* 6 *October* 1938)

1. The problem

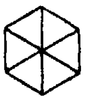
The structure of the parent of the aromatic series has long been one of the foremost unsolved problems of chemistry. The following formulae will be a sufficient reminder of the long, complicated period of historical development, which extended from 1865 until the present decade:



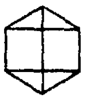
Kekulé



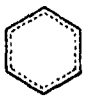
Dewar



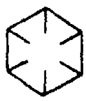
Claus



Ladenburg



Thiele



Baeyer

The fundamental difficulty for chemists has always been that of reconciling the stability of the benzene nucleus with its chemical transformations. If one considers only the transformations, Kekulé's expression is undoubtedly the best single representation. Dewar's is a possible addition if a dynamical system be postulated. But neither structure interprets the stability of the ring; nor do the saturated, but highly strained alternatives of Claus and Ladenburg. Formulae which employ imperfectly defined symbols leave us with the problem of more exactly defining the significance of symbols. As to dynamical systems, many have been suggested; but none escapes the difficulty that the stability of a system of tautomerides cannot be greater than that of its most stable component: the whole trouble is to invent a sufficiently stable component. Moreover, we now know that there is a stereochemical criterion to be satisfied: no formula, static or dynamic, is satisfactory which does not provide the ring with the at least approximate planarity demanded by the X-ray analysis of aromatic crystals, and by electron diffraction in benzene vapour. I shall deal in turn with the modern theory and with some recent experimental evidence.

2. AN ATTEMPTED THEORETICAL SOLUTION

The modern theory of benzene structure is not simply one more of the *ad hoc* hypotheses which have from time to time been advanced in order to provide some symbol which could be supposed to express the peculiarities of benzene. Our modern views on benzene arise in a natural way as a part of a much more extensive theoretical development. The generic relationship between the branches of theory here referred to is the following. First, the electronic theory of valency, in its original form, covers in a satisfactory way a large part of the general theory of chemical combination. Secondly, an extension to the original electronic theory of valency (I shall call the extension the theory of mesomerism) has been created in order to deal with another large part of the phenomena of chemical combination for which the original theory of valency is inadequate. The theory of benzene structure is a "rider", because in this general extension to the theory of valency benzene theory is automatically included as a special (and especially important) case. Fundamental to every link in the chain of theory is the quantum theory. The relationship may be outlined as follows.

Commencing with the quantum theory, we know that it is impossible accurately to specify the position and motion of an electron, considered as a particle, at each instant of time. What can be known about an electron, considered as a particle, is of a statistical nature, and is summarized in the behaviour of a wave. The frequency of the wave measures the energy of the electron ($E = h\nu$) and the amplitude of the wave measures the probability that the electron is in a given element of volume (prob. $= |\psi|^2 dv$). The fundamental wave-property in the theory of the *atom* is *self-interference*—the self-interference of a confined wave; this annihilates all frequencies except those that can form standing waves, which are perpetuated and describe the quantum states. The fundamental wave property in the theory of *valency* is *resonance*—the resonance of connected standing waves; their mutual perturbation replaces these waves by new standing waves, one of which usually corresponds to the creation or modification of some bond.

In order to appreciate the significance of resonance in the theory of valency, it is useful to recall the nature of resonance in vibrating mechanical systems. The simplest example is that of two like pendulums hung from the same stretched string. If one pendulum is given a displacement, its interaction with the string causes the vibratory motion gradually to become transferred to the second pendulum, after which it travels back to the first as is shown in the right-hand time graphs of Fig. 1. This motion is obviously not simple harmonic, and in general it is not even singly periodic, *for at no time after the start is the original configuration exactly* repeated. It can, however, be shown to be a mixture of two simple harmonic motions. One of these occurs alone if we start the pendulums with equal amplitudes in phase; the frequency $\nu - \epsilon$ is then a little less than the frequency ν of either uncoupled pendulum, because the two pendulums when swinging in phase deform the supporting string in mutually accommodating ways and this lowers the restoring forces. The other simple harmonic motion occurs when we start the pendulums with equal amplitudes in opposite phase; the frequency $\nu + \epsilon$ is *now* a little *greater* than that of the uncoupled pendulums because of the opposing actions on the string. When we start the pendulums with any arbitrary amplitudes and phase difference, the two simple harmonic motions occur together, and the "beats" arising from their differing frequencies give rise to the complicated motion described. However, it is less the beats than the analysis into two simple harmonic motions which is important from the point of view of analogy.

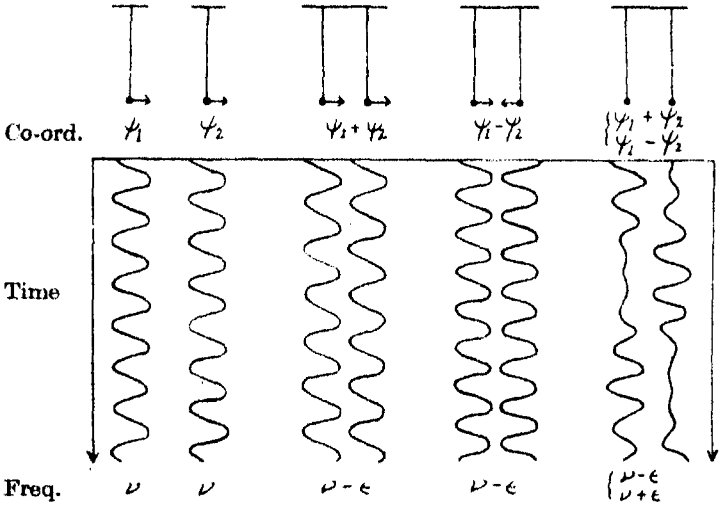

Co-ord.
ψ_1
ψ_2
$\psi_1 + \psi_2$
$\psi_1 - \psi_2$
$\begin{cases}\psi_1 + \psi_2 \\ \psi_1 - \psi_2\end{cases}$
Time
Freq.
ν
ν
$\nu - \epsilon$
$\nu - \epsilon$
$\begin{cases}\nu - \epsilon \\ \nu + \epsilon\end{cases}$

FIG. 1. Mechanical resonance: time graphs illustrating the behaviour of uncoupled and coupled pendulums. (*Reproduced from* Nature *by permission.*)

If ψ_1 and ψ_2 are the co-ordinates measuring the displacements of the two pendulums, then before coupling, each co-ordinate was associated with a definite frequency ν, but after coupling neither ψ_1 nor ψ_2 has any definite, characteristic frequency. On the other hand, the linear combination $\psi_1+\psi_2$ has the constant, characteristic frequency $\nu-\epsilon$, whilst the combination $\psi_1-\psi_2$ has an equally constant and characteristic frequency $\nu+\epsilon$. We may say that the mutual perturbation, or "resonance", has destroyed the former co-ordinates of periodic motion, ψ_1 and ψ_2, and has set up in their place two new co-ordinates each with definite periodicity, these new co-ordinates, $\psi_1+\psi_2$ and $\psi_1-\psi_2$, being independent linear combinations of the old ones. At the same time the original co-incident frequency, ν, has been split into two new frequencies, one greater and one less than the original.

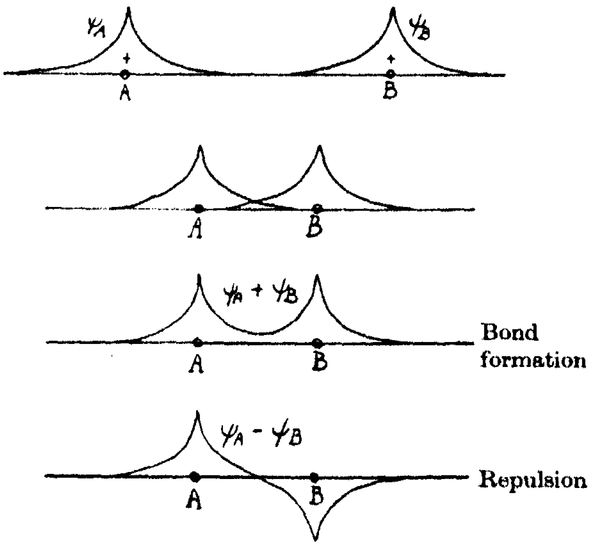


FIG. 2. Quantum mechanical resonance: schematic representation of the formation of a bond. (*Reproduced from* Nature *by permission.*)

If there are alternative ways in which we can supply electrons to a fixed system of atomic nuclei, and if the standing waves representing these alternatives can perturb each other then there will be "resonance": it is called "resonance" on account of the analogy with mechanical resonance, though the analogy has certain limitations. Consider the imaginatively simplified example, represented in fig. 2, of two atoms in a one-dimensional world, each atom having only one electron. So long as the atoms are well separated the electron in atom A will move according to the standing wave ψ_A, whilst that in atom B will move according to the standing wave ψ_B. The frequency of each wave corresponds to the energy of the atom, and the form of the wave determines the probability, or charge distribution. If now the nuclei are moved closer together so that the fringes of ψ_A and ψ_B overlap, each wave will perturb the other and neither will any longer be a standing wave of characteristic frequency. Two new standing waves will arise; one, the like-phase combination, $\psi_A+\psi_B$, which has a decreased frequency, and the other, the opposite-phase combination, $\psi_A-\psi_B$, which has an increased frequency. The former represents a system of reduced energy, that is, a more "stable" system—one which will tend to be formed, and to persist; clearly it corresponds to the establishment of a "bond".

This is a wave picture: a particle picture of the same phenomenon may be drawn as follows. It is a direct consequence of the uncertainty principle that all particles whose motion is restricted have vibrational energy. The principle may be expressed by the formula $\Delta q . \Delta p \sim h$, where Δq represents the uncertainty of position of the particle, Δp the uncertainty in its momentum, and h is Planck's constant; the momentum is closely related to the kinetic energy ($T=p^2/2m$) and therefore to the minimum possible value of the total energy, E. A particle which is confined to a very small region of space ($\Delta q\rightarrow 0$) would necessarily be moving with a correspondingly large energy ($\Delta p\rightarrow\infty$; $E\rightarrow\infty$); and in proportion as the spatial region available to the particle is expanded the energy of the particle will be reduced; wherefore the system which contains the particle, having less energy, will be more stable. Thus if a system containing electrons can find a way of changing such that new motions become permitted to the electrons, the system will in fact undergo that change, becoming more stable. Consider again the two one-dimensional one-electron atoms. So long as they are well separated, the electron in each is restricted to move in the neighbourhood of its nucleus with a position probability defined by the curve called the wave function. But if, notwithstanding the mutual repulsions of the nuclei, we bring them into such positions that the regions occupied by the separate curves overlap, then each electron will be able to escape out of the domain of its own nucleus into that of the other: the electrons can change places. This newly acquired freedom of motion will give added stability to the system, wherefore it will tend to persist, and we shall describe this by saying that a bond has been formed.

In the atoms and molecules of the real, three-dimensional world other restrictions, of which the chief is Pauli's principle, apply to the motions of electrons, but this does not alter the general nature of the basis on which chemical combination is interpreted. Historically, the two most important events in the development of the electronic theory of valency in its original form are, first, the purely empirical recognition of the two-electron bond by Lewis in 1916, and, secondly, its physical interpretation by Heitler and London, in 1928, on the general basis indicated.

We now come to the extension of valency theory summarized in the term mesomerism. This extension was constructed to deal with the very large class of molecules, including all aromatic molecules and many others, for which ordinary valency theory gives a wrong picture of the electron distributions. We meet with the same two evolutionary steps. First, it was empirically recognized that the valency picture is wrong, whenever, starting with an electron distribution given by ordinary valency theory, it is possible, with nuclei in fixed relative positions, to displace the electron-pairs in such a way as to produce a second valency structure, which, like the first, obeys the valency rules. In all such cases, some part of that possible electron displacement is permanently present in every molecule. In formamide, for example, some part of the electron displacement which would convert $NH_2.CH:O$ into $NH_2:CH.O$ is present permanently in every formamide molecule, so that the actual electron distribution is a mean between those which might have been assumed on the basis of ordinary valency theory. I have since 1926 derived this view from the facts of organic chemistry including many which Flürscheim had collected together and had attempted to explain without assumptions as to the physical nature of forces responsible for valency. Especially important amongst these facts were those relating to chemical equilibrium, since equilibria depend on the normal, non-activated states of molecular systems.

The second step, the physical interpretation of the phenomenon, was developed independently by Pauling and me between 1929 and 1933. The central principle employed is the same as that used in the interpretation of bond formation. It is the uncertainty principle, or, alternatively, the derivative principle of resonance, which is fundamental to all departments of the theory of atomic and molecular interaction, including even van der Waal's attraction and steric repulsion. In its application to mesomerism the qualitative conception may be expressed as follows: molecules depart from the normal valency arrangements of the electrons for the same reason that atoms depart from the electronic arrangements characteristic of free atoms, namely, in order to produce a more stable system; and the molecules achieve the added stability in essentially the same way as atoms do, namely, by discovering new possibilities of motion for their electrons.

Let us first consider the wave picture. We can regard an ordinary structural formula as a symbol of the standing wave which has been derived by taking into account all the resonance involved in the bonds of the formula and selecting from the final linear combinations of wave functions that which has lowest energy. Mesomerism arises when, with nuclei in fixed positions, we can write several ordinary structural formulae; for in this case the corresponding "final" combination waves are not really final. They have to be linearly combined once more, with a further delocalization of the electrons, a further splitting of frequencies, and therefore a further gain in the stability of the most stable state which is possible for the system. It is essentially as a matter of convenience rather than principle that we separate this final stage of approximation to a true

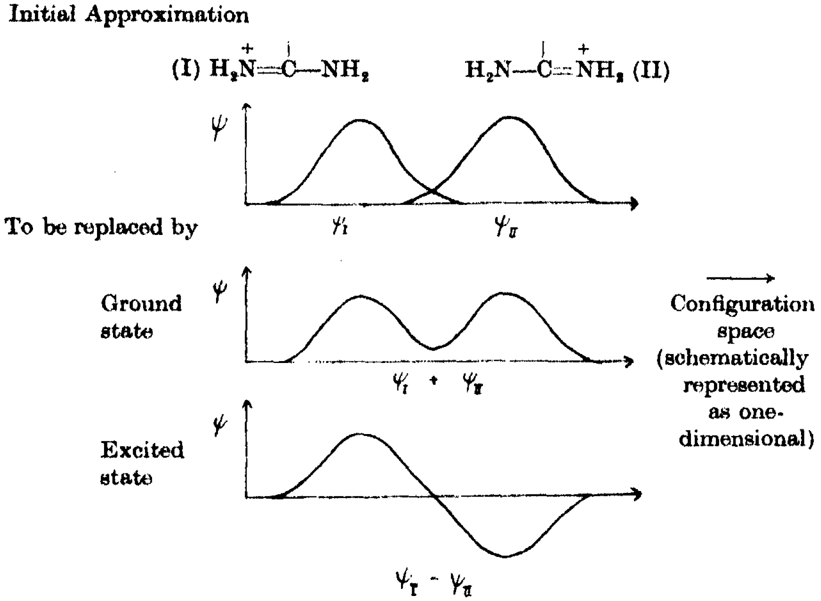

Initial Approximation
(I) $H_2\overset{+}{N}{=}\overset{|}{C}{-}NH_2$
$H_2N{-}\overset{|}{C}{=}\overset{+}{N}H_2$ (II)
ψ
ψ_I
ψ_{II}
To be replaced by
Ground state
ψ
$\psi_I + \psi_{II}$
Configuration space (schematically represented as one-dimensional)
Excited state
ψ
$\psi_I - \psi_{II}$

FIG. 3. Quantum mechanical resonance: schematic representation of mesomerism in the amidinium ion. (*Reproduced from* Nature *by permission.*)

description from the previous stages, these being summarized in the classical structural ideas which the conception of mesomerism attempts to improve.

As an example we may consider the amidinium ion. The standing waves corresponding to the alternative formulae are similar; strictly expressed they are polydimensional, but are represented in fig. 3 schematically in one dimension. The final combination wave is a mixture of equal proportions of the two, and hence the charge distribution may be described as intermediate between those corresponding to the alternative formulae given by ordinary valency theory. The identical (fictitious) energy levels represented by the alternative formulae are, as usual, split into a ground level which lies lower, and an excited level which lies higher, than the original coincident levels, as is shown in fig. 4. We usually speak of the ground state as "the" mesomeric state, and the energy interval between this state and that of the formulae given by elementary valency theory as the mesomeric energy.

When the alternative formulae are dissimilar, as in the amide group, there will be upper and lower energy levels belonging to the formulae of elementary valency theory, and the corresponding waves will not mix together in equal proportions. The combination wave of the ground state will contain an excess of the lower initial approximation wave and that of the excited state an excess of the upper (fig. 4). But still the real ground level will lie below the lower of the initial, fictitious levels, and the charge distribution will have a certain "intermediate" character between those of the elementary valency structures.

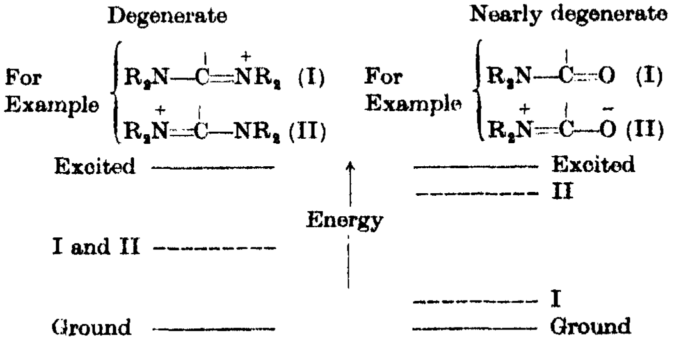

Degenerate
Nearly degenerate
For Example
$R_2N—\overset{|}{C}{=}\overset{+}{N}R_2$ (I)
$R_2\overset{+}{N}{=}\overset{|}{C}—NR_2$ (II)
For Example
$R_2N—\overset{|}{C}{=}O$ (I)
$R_2\overset{+}{N}{=}\overset{|}{C}—\overset{-}{O}$ (II)
Excited
I and II
Ground
Energy
Excited
II
I
Ground

FIG. 4. Splitting of energy levels.

In the corresponding particle picture the essential point is the greater delocalization of the electrons, and their greater opportunities of motion. This produces a smaller range of variation of the kinetic energy, and a smaller total energy, in the mesomeric state than in those fictitious states which are given by classical valency theory. We may again clarify the conception slightly by proceeding to the limit of imaginative simplification: consider a one-dimensional molecule with one *displaceable* electron. Suppose that in one structure suggested by an elementary theory appropriate to the one-dimensional case the electron is represented as being confined to the left-hand half of such a molecule, whilst in another it is represented as confined to the right-hand half, as illustrated in fig. 5. The two structures suggest an analogy with the two positions for the double bond in the elementary representations of (e.g.) formamide. For each structure there will be a curve, localized to the appropriate half of the linear molecule, which defines the probability of position of the electron, considered now as a particle. However, the real molecule will be one in which the particle electron has the power of escape from one half of the molecule to the other, and it is that power of escape which gives the added stability.

When mesomerism was first empirically recognized as a general phenomenon of chemistry the aspect which was most strongly emphasized was that of electron redistribution. At that time very little was known about dipole moments, and it was thus possible to predict in 1926 that evidence of the electron redistribution required by theory would be found in dipole measurements. The first evidence, which related, however, only to a few special cases, emerged from Hojendahl's work published in 1928, whilst the first comprehensive investigation, definitely designed to reveal the expected effect, was published by Sutton in 1931. He showed that, where the conditions for mesomerism exist in a compound that has a finite dipole moment, then that moment shows an abnormality in the expected direction.

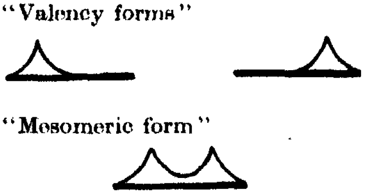


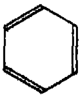
FIG. 5. *Mesomerism: simplified model of one-dimensional molecule with one displaceable electron.*

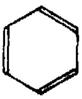
The complementary aspect of mesomerism, that of the energy disappearance, was established qualitatively by Burton and me in 1929 as the cause of abnormal molecular stability, often very strikingly exhibited, which older theories were unable to explain. But this additional stability was first given a satisfactory basis in measured energy values, and was first shown to be general amongst the structures capable of mesomerism, by the work of Pauling and Sherman in 1933. They collected together a large body of thermochemical data, which showed that, in all those molecules in which this special stability was expected, it could be detected as an abnormally large heat of formation, the deviation amounting sometimes to several electron-volts.

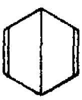
There are two other general effects of mesomerism, and they have a close causal connexion with those already mentioned. It is an immediate consequence of the electron redistribution that the differences between what elementary valency theory would regard as single and double bonds become always reduced, and in some cases eliminated. Thus in formamide neither is the N—C bond single nor is the C—O bond double: both are intermediate between single and double. Such bonds are found to be shorter than single and longer than double bonds; they also exhibit a rigidity intermediate between those of single and double bonds. When the difference between adjacent bonds is completely eliminated, additional molecular symmetry may result, as in the carbonate ion, the three bonds of which used to be regarded two as single bonds and one as double bond. We now know from spectroscopic evidence (derived mainly by Schaeffer) that, as theory requires, all three bonds are identical—identical geometrically, mechanically and electrically, having the same length, the same force constant, and the same electric moment.

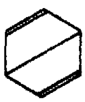
The fourth general effect of mesomerism is related directly to the enhanced stability of mesomeric systems, for this, as we have seen, is associated with more widely dispersed electron orbits. The expected greater latitude in the electron motions can in general be detected as either an increased refractive index, or an increased diamagnetic susceptibility, or both, although, for reasons which are qualitatively understood, there is not a simple parallel between these effects, and those found in the dipole moment, the energy content, or the bond constants.

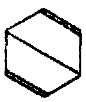
In the example of benzene the theory of mesomerism has been set out in an elegant, semi-quantitative manner by Pauling and Wheland. First, they showed that the following five formulae, constitute a *complete set* of the valency-theory electron distributions, which, by becoming mixed together, can produce less localized and therefore more stable electron distributions.


I


II


III


IV


V

$$\psi_{(\text{Benzene})} = 0{\cdot}6216(\psi_{\text{I}} + \psi_{\text{II}}) + 0{\cdot}2711(\psi_{\text{III}} + \psi_{\text{IV}} + \psi_{\text{V}}).$$

Secondly, they worked out the mixing proportions (shown in the coefficients attached to the wave functions under the formulae) which would give the most stable distribution theoretically possible—the distribution which we must seek to identify with the actual one.

A brief reference may be made here to the appearance or otherwise in benzene of the four general effects of mesomerism. First, the dipole moment is necessarily zero, so that no observable effect of the electron redistribution on the dipole moment can be expected. Secondly, the excessive stability of benzene has been qualitatively known since the time of Kekulé; and now Pauling and Sherman have estimated the energy defect as 1·7 eV from thermochemical values. Pauling and Wheland could not calculate this quantity from their theory in an absolute manner; but they did calculate the ratio of the energy defects in benzene and naphthalene, obtaining the value 1 : 1·9, which agrees to this degree of accuracy with the value estimated thermochemically. Thirdly, with regard to the molecular geometry of benzene, it is a consequence of the equality in the proportions in which the two Kekulé, and likewise the three Dewar, structures, enter into the mixed wave function for benzene, that the molecule should have the full symmetry of a plane regular hexagon. The experimental evidence on this subject is discussed in the next section. Fourthly, evidence of electron delocalization emerges clearly from the study of the diamagnetic susceptibility of benzene derivatives. Such studies have been made by Krishnan and Lonsdale and the data have been interpreted by Pauling and Lonsdale. The contribution of an electron to diamagnetic susceptibility depends upon the area which can be circumscribed by the electron in a plane perpendicular to the measuring field. Most valency electrons yield a susceptibility contribution corresponding to an area of the order of an atomic cross-section, and this is true also of benzene derivatives for magnetic fields applied in the plane of the ring. When, however, the field is applied perpendicularly to the ring, an increased diamagnetic effect is observed, which can be accounted for if we suppose that six of the electrons can freely circumscribe the whole ring. Thus it appears that the ring is a free conductor with respect to the six electrons, which elementary valency theory would have allocated to the three bonds, now regarded as delocalized.

3. An attempted experimental solution

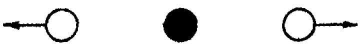
There are two types of method available for the experimental determination of any simple molecular model; the first consists in X-ray or electron interferometry; the second may be summarized under the name "long-wave spectroscopy". The sensitivity and scope of the interferometric method is limited for hydrocarbons, but it has shown in the case of benzene that this molecule conforms approximately to the model indicated as most probable by theory. On the other hand, the spectroscopic method, which is sensitive, exact and particularly suitable for application to the case of benzene, has been claimed to give quite contrary results. In order to explain this, it will be necessary to indicate some of the more important principles underlying the spectroscopic method.

In long-wave spectroscopy one studies the internal vibrations of molecules with the object of deducing, first, the configuration of the atoms, and secondly, the force field which maintains that configuration. The vibrations of a molecule, as of any mechanical system, consist in a superposition of harmonic vibrations of which the chief are a set of fundamentals, equal in number to the degrees of freedom of internal atomic motion. Any harmonic vibration may record its frequency in a spectrum, provided the physical mechanism of the production of the spectrum is suitable. This is not always the case.

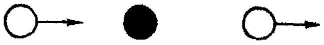
One of the spectra used is that of infra-red absorption. The mechanism of its production may be explained as follows. When a light beam travels past a molecule, the electric vector of the light causes a periodic surge of the electrons relatively to the nuclei, and therefore a periodic variation of the dipole moment. If the molecule has a natural vibration *which involves a periodic variation of the dipole moment,* and if the frequency of the light happens to agree with this natural frequency, then the light will strongly excite these particular vibrations, losing energy in the process, so that we shall observe an absorption band. It follows that only those vibrations which involve an oscillation of the dipole moment will record themselves in the infra-red spectrum.

Another important spectrum is the Raman, or scattering spectrum. The following is a simple picture of the mechanism of its production. Visible or near-ultra-violet light sets up (comparatively rapid) forced dipole oscillations, which radiate secondary light, that is, the scattered light. Normally the secondary light has the frequency of the dipole oscillations, and therefore of the incident light; that, at least, is true if the molecule is stationary. If, however, the molecule is executing a (comparatively slow) natural vibration *which involves a periodic variation of the internal electrical elasticity,* then the amplitude of the forced dipole oscillations will fluctuate correspondingly; and so also will the amplitude of their radiation. Thus the natural molecular frequency will appear in the scattered radiation as a "beat" frequency, and it is this "beat" which (in the usual way as a difference of two harmonic terms) the Raman spectrum records. It follows that only those vibrations which are associated with a periodically varying electrical elasticity will leave their record in the Raman spectrum.

The original spectroscopic conclusion regarding benzene was that it could not have a centre of symmetry. Concerning the presence or absence of centres of symmetry, a simple rule, of the type called a "selection rule", follows directly from the two physical pictures given. This may be illustrated by means of the two linear vibrations of carbon dioxide, which has a centre of symmetry. The motions of the oxygen atoms relatively to the carbon atom (it is simpler here to use the carbon atom rather than the centre of gravity as a reference point) are shown on fig. 6. In the symmetric vibration the centres of gravity of both nuclei and electrons remain fixed at the centre of symmetry throughout the vibration; and hence the dipole moment does not vary. Therefore the symmetric vibration cannot record its frequency in the infra-red spectrum. In the antisymmetric vibration the effect at any moment of the displacement of one oxygen atom on forces holding the electrons will be counterbalanced by the simultaneous displacement of the other; so that the overall electrical elasticity will not vary. Therefore the anti-symmetric vibration will not appear in the Raman spectrum. This principle is quite general. If a centre of symmetry is present, all harmonic vibrations are either symmetric or antisymmetric to it; and, since those that are symmetric are excluded from the infra-red whilst those that are antisymmetric are excluded from the Raman spectrum, every vibration must be excluded from one spectrum or the other. There may be special reasons why some vibrations are excluded from both, but all vibrations must be excluded from one. Therefore if we in fact find the same vibrations appearing in both spectra the conclusion cannot be escaped that there is no centre of symmetry. This was the difficulty with benzene, the older records showing as many as twelve coincident frequencies in its infra-red and Raman spectra; it appeared, therefore, that the plane, regular hexagonal model, indicated as most probable by the present position of the theory, could not be correct.



Symmetric vibration: dipole moment constant (= 0): frequency absent from infra-red.



Antisymmetric vibration: polarizability constant: frequency absent from Raman

FIG. 6. Simplified representation of the symmetric and antisymmetric linear vibrations of carbon dioxide.

My colleagues, Dr W. R. Angus, Mr C. R. Bailey and Dr C. L. Wilson, together with several of our co-workers, have recently taken up this question. We set ourselves three problems. The first was the limited one of finding out the meaning of these coincident frequencies. The next was the more general problem of empirically establishing (without reference to theory) the correct model for benzene, and placing it on an unassailable experimental foundation. The final, and most difficult task (it is not yet completed), is to define the force-field which maintains this model. In each of these stages use is made of the heavy isotope of hydrogen.

In the problem of the coincidences the hydrogen isotopes can be applied to determine whether two identical frequencies in the infra-red and Raman spectra really belong to the same vibration, or whether they belong to *different vibrations which by chance have the same frequency. We can examine both* C_6H_6 *and* C_6D_6*: if the vibration is the same then the frequency* shifts will be the same, and what was a coincidence in benzene will remain a coincidence (at a different frequency) in heavy benzene. But if the vibrations are different, then the laws which govern the frequency shifts will be different, and what was a coincidence in benzene will no longer be one in heavy benzene. In conducting the experiments it transpired, most surprisingly in view of the literature, that there were no coincidences requiring this test. *Fig.* 7 shows a plot of the low-frequency half of the Raman spectrum and of the infra-red spectrum of benzene vapour. The toleration to be allowed is at most 4 cm.$^{-1}$, and within this limit there are no coincidences. The same is true for C_6D_6, for which the corresponding plots are in Fig. 8.

The reason why coincidences had been reported became clear when the spectra given by the liquids were examined. A number of additional frequencies were then observed and it was amongst these that the coincidences appeared. Fig. 9 shows a plot of the low frequency half of the spectra for liquid benzene: a number of coincidences can be seen. Quite similar results were obtained for liquid C_6D_6, the corresponding diagram for which is in fig. 10. It is true that infra-red spectra like that of fig. 9 have been recorded as the spectrum of benzene vapour, but there can be no doubt that this is because thin films of liquid have been allowed to condense on the end-plates of the absorption tube. When we avoid this occurrence by means of an appropriate experimental arrangement the spectrum is of the simple form shown in fig. 7. The reason why the liquids show all the extra bands which give the coincidences is as follows: in the liquid the molecules are under intermolecular forces, which, though fluctuating, are at every moment very slightly deforming the equilibrium configuration about which the molecule vibrates, thus destroying the strict centre of symmetry, which the results for the vapour show the undisturbed molecule to possess. We have here an illustration of how sensitive

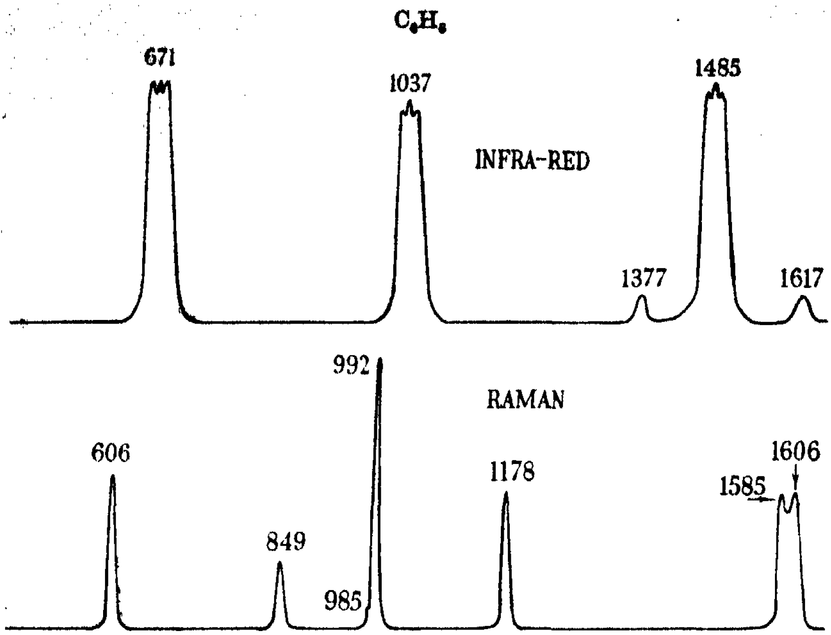


FIG. 7. Raman and infra-red spectra of benzene (vapour).

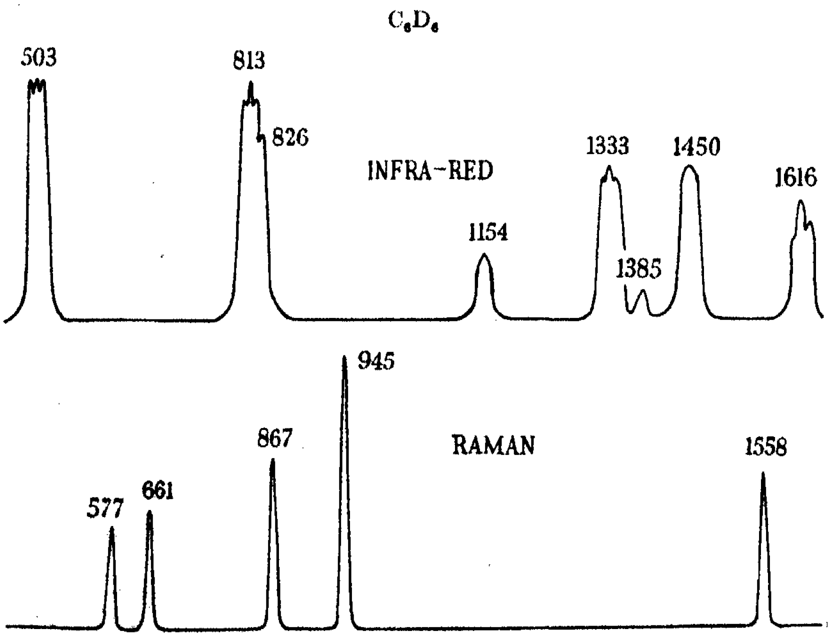


FIG. 8. Raman and infra-red spectra of hexadeuterobenzene (vapour).

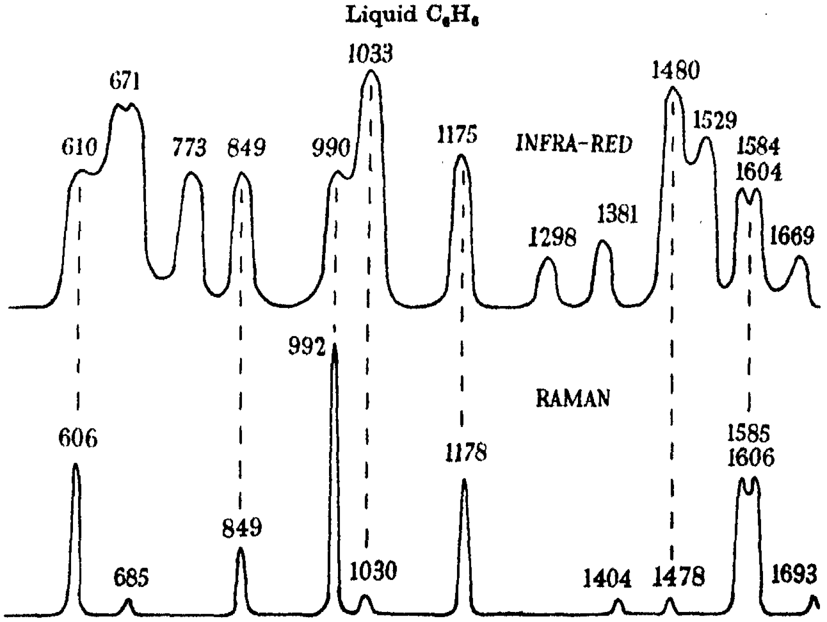


FIG. 9. Raman and infra-red spectra of liquid benzene.

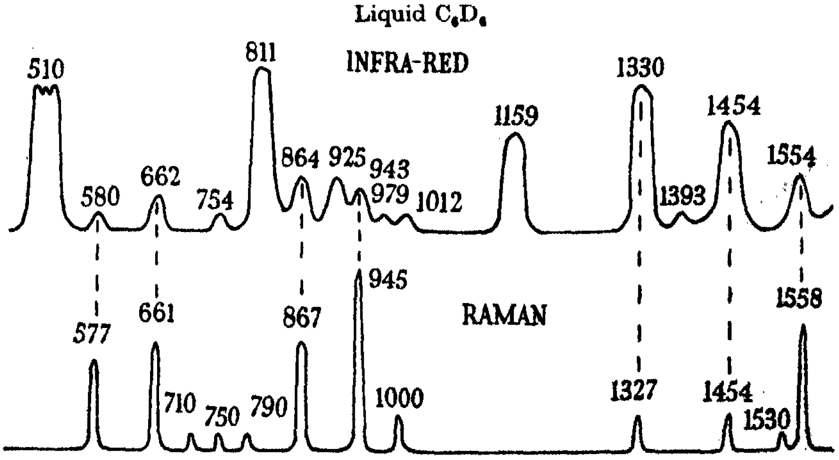


Fig. 10. Raman and infra-red spectra of liquid hexadeuterobenzene.

is the spectroscopic test for symmetry. It is also a typical example of the way in which selection rules in general may be used to determine what elements of symmetry are present in any molecule.

As a check we may observe that when the centre of symmetry of benzene is deliberately destroyed by a suitable form of partial deuterium substitution, as in 1:3:5-$C_6H_3D_3$, then coincidences appear in the spectra of the vapour. Fig. 11 shows a plot of the low frequency end of the infra-red and Raman spectra of this substance, and one can observe a number of coincidences. The whole spectrum shows seven coincidences of fundamentals.

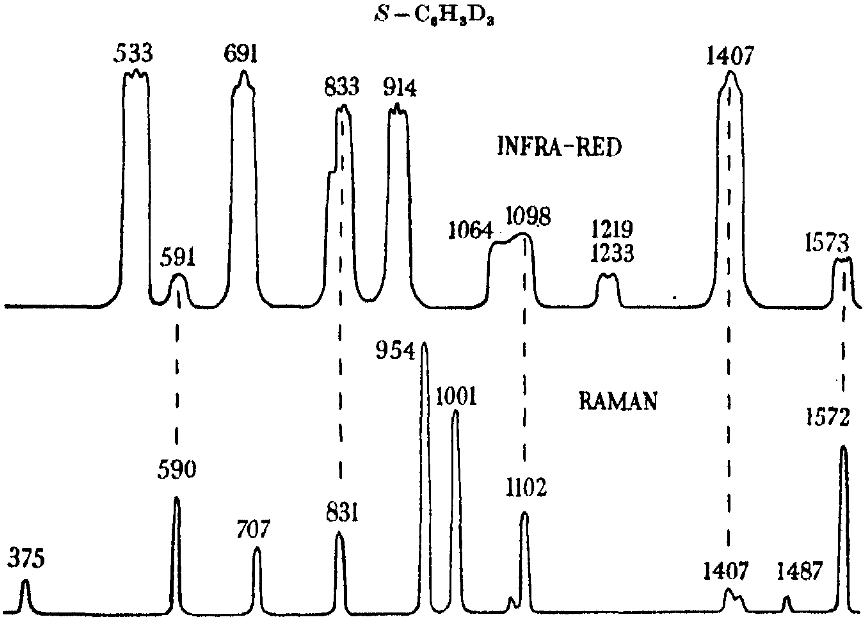


Fig. 11. Raman and infra-red spectra of 1:3:5-trideuterobenzene (vapour).

The only known example of a genuinely accidental coincidence of fundamentals is in the spectrum of 1:4-$C_6H_4D_2$, which, having a centre of symmetry, ought not to show any coincidences. It shows, as a matter of fact, one, but the coinciding frequencies (597 cm.$^{-1}$ in the Raman spectrum and 595 cm.$^{-1}$ in the infra-red) belong to quite different vibrations as can be proved by comparison with other benzenes, e.g. 1:2:4:5-$C_6H_2D_4$, which has the same symmetry.

It is necessary now to consider the second problem, the establishment of the correct model. Our ability to modify benzene by isotopic substitution is here used in accordance with the following principle. When we replace atoms by their isotopes, all the forces within the molecule remain unchanged, because the forces depend on the interactions of electric charges which are not altered in an isotopic substitution. Therefore the frequency shifts which we observe will be due, not to any change of forces, but solely to known changes in the masses of the atoms. Hence, *for an assumed model*, we can calculate relationships between the old frequencies and the new. Since we can do this, not only for C_6H_6 and C_6D_6, but also for the eleven other benzenes obtainable by the partial replacement of protium by deuterium, the number of verifiable frequency relations is large. A model that satisfies, not only all the selection rules (which fix the symmetry), but also all these quantitative relations, as well as explaining the numerous details of the spectra such as intensities, infra-red band forms, Raman polarizations, and the occurrences of combination tones, may be considered proved.

Strictly, we should now consider all possible models, but I shall simplify discussion here by anticipating the establishment, which is now quite complete, of the plane, regular hexagonal model. This model has twenty fundamental frequencies whose vibration-forms are shown in fig. 12. The arrows indicate motion in the plane of the ring, and the noughts and crosses motions perpendicular to the plane. The twenty vibration forms are divided by the horizontal and vertical lines into ten classes according to their symmetry. Thus, the first row of vibrations have sixfold symmetry with respect to the unique axis, the second threefold symmetry, the third row twofold symmetry, and the last no symmetry, all with respect to the unique axis. The further subdivisions (indicated by the vertical dividing lines) depend on differences with respect to other elements of symmetry, such as the centre, the plane, and the twofold axes in the plane of the ring. It is unnecessary to enter into all these details, but it may be noted here that the first two vibrations have all the symmetry of the model, while all other eighteen vibrations lack some part of the total symmetry.

By considering, just as for carbon dioxide, how pairs of equivalent atoms either compensate, or fail to compensate, each other's effect on the variation of, first, the dipole moment, and second, the electrical elasticity, we can decide in which spectra each of these vibrations is able to record its frequency. Of the twenty frequencies, seven (labelled *R*) should show in the Raman effect. For reasons into which it is unnecessary to enter, the two vibrations which have all the symmetry of the model should give polarized Raman lines, whilst all the others should give depolarized lines. Of the remaining vibrations, four (labelled *I*) should appear in the infra-red spectrum. The one which is alone in its symmetry class should give what is called a "parallel" band, and the other three "perpendicular" bands. The remaining nine vibrations cannot record their frequencies either in the Raman or in the infra-red spectrum.

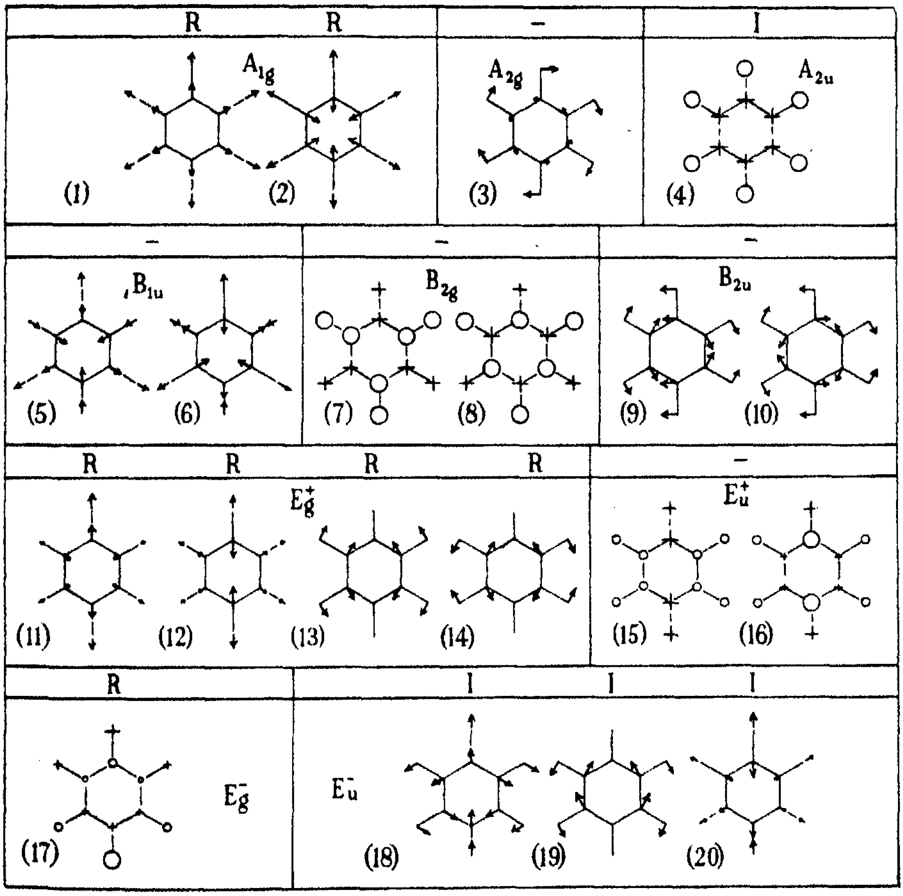


FIG. 12. Vibration forms of benzene (D_{6h}) model.

NOTES: (1) Vectors perpendicular to plane of ring represented by + ○.
(2) For each type of atom ↑: ↑ = + : + = ○ : o = 2 : 1.

It is possible to decide to which vibrations the observed frequencies belong, partly, and in a few cases, by purely spectroscopic characteristics such as polarization, but more generally by means of the isotopic frequency shifts. The principles of the latter method are contained in a theorem discovered first by Teller, and, independently, by Redlich.

The Raman spectra of C_6H_6 and C_6D_6 are shown diagrammatically in fig. 13. The expected seven fundamental frequencies appear each as a well-defined line in the spectrum of C_6D_6, and the same is true of C_6H_6, except that, for an understood but purely incidental cause (resonance arising from an accidental agreement of frequency between a fundamental and a combination tone) one of the fundamentals, that at 1596 cm.$^{-1}$, appears as a doublet. The identification of these frequencies with the Raman-active vibrations is as follows. The polarized line at 992 cm.$^{-1}$ frequency displacement in C_6H_6, and its counterpart at 945 cm.$^{-1}$ in C_6D_6, represent the "breathing vibration" of the carbon ring (fig. 12, formula 1): the six carbon atoms, each carrying its hydrogen with it, pulsate inwards and outwards synchronously.* The vibration is only slightly slower in C_6D_6 than in C_6H_6 because the moving unit is essentially the whole CH group which changes in mass by the small ratio 13:14. The highest frequency (polarized) line, at 3062 cm.$^{-1}$ in C_6H_6 and 2292 cm.$^{-1}$ in C_6D_6. is the breathing vibration of the hydrogen atoms (fig. 12, formula 2): it represents the rapid synchronous pulsation of the hydrogen atoms with only a very little related motion in the carbon atoms; and, of course, it is superposed on the slower breathing movement of the CH groups as a whole. Here the frequency shift is large because almost all the motion is in the hydrogen atoms, whose mass changes in the large ratio 1:2. The line at 849 cm.$^{-1}$ in C_6H_6 and 661 cm.$^{-1}$ in C_6D_6 is identified by its frequency shift with the vibration in which the plane of the hydrogen atoms rocks over the plane of the carbon ring about an axis common to both planes (fig. 12, formula 17). The low frequency line at 606 cm.$^{-1}$ in C_6H_6 and 577 cm.$^{-1}$ in C_6D_6 represents the vibration in which the ring is longitudinally stretched and compressed with bending of the ring angles (fig. 12, formula 11). The higher frequency at 1596 cm.$^{-1}$ in C_6H_6 and 1558 cm.$^{-1}$

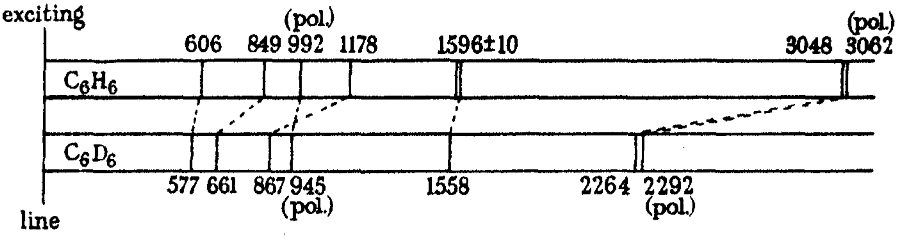


FIG. 13. Raman spectra of C_6H_6 and C_6D_6: correlation of fundamentals.

---

* Working models were shown on this and of all the vibrations hereafter mentioned.

in C_6D_6 corresponds to a similar longitudinal extension and contraction effected mainly by alteration of the carbon-ring bond lengths (fig. 12, formula 13). In both these vibrations of the carbon ring the frequency shift is small, because the carbon atoms carry their hydrogen atoms with them. The remaining high frequency, 3048 cm.$^{-1}$ in C_6H_6 and 2264 cm.$^{-1}$ in C_6D_6, is a bond-stretching frequency of the hydrogen atoms (fig. 12, formula 12), and the lower frequency, 1178 cm.$^{-1}$ in C_6H_6 and 867 cm.$^{-1}$ in C_6D_6, is a bond-bending frequency of the hydrogen atoms (fig. 12, formula 14), these two vibrations having the same kind of symmetry as the two preceding. The frequency shifts of the hydrogen atom vibrations are large because of the proportionately large mass change of the moving unit.

The frequencies of four more fundamentals can be found from the infra-red spectra, the principal bands in which are represented diagrammatically in fig. 14. The frequency shifts show that the expected four active fundamentals are represented in C_6H_6 by the four bands of outstanding intensity; and in C_6D_6, as to three by the three very strong bands, and as to the fourth by the moderately intense band at 1333 cm.$^{-1}$. All other bands are combination tones. The low frequency, 671 cm.$^{-1}$ in C_6H_6 and 503 cm.$^{-1}$ in C_6D_6, represents the vibration in which the six hydrogen atoms together move perpendicularly to the plane of the ring (fig. 12, formula 4). The frequency at 1485 cm.$^{-1}$ in C_6H_6 and 1333 cm.$^{-1}$ in C_6D_6 corresponds to the vibration (fig. 12, formula 18) in which the ring is deformed as if to outline a conventional heart. The high frequency, 3080 cm.$^{-1}$ in C_6H_6 and 2294 cm.$^{-2}$ in C_6D_6 (in C_6H_6 there is doublet formation due to resonance with a combination tone) belongs to a hydrogen-bond stretching vibration having the same type of symmetry (fig. 12, formula 20). Finally the frequency at 1037 cm.$^{-1}$ in C_6H_6 and 813 cm.$^{-1}$ in C_6D_6 corresponds to a hydrogen-bond bending vibration (fig. 12, formula 19) of the same symmetry as the two preceding vibrations.

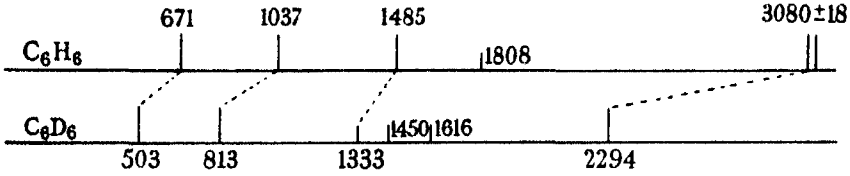


FIG. 14. Infra-red spectra of C_6H_6 and C_6D_6: correlation of fundamentals.

Thus we establish the frequencies of eleven of the vibrations: it is considerably more difficult to deal with the remaining nine. Two methods have been employed with some success.

One consists in examining the ultra-violet emission (fluorescence) spectra which arise when the electronically excited benzene molecule regenerates the ordinary (but vibrating) benzene molecule. As Kistiakowsky first recognized, two types of spectra can be thus obtained. One is formed if the electronically excited benzene molecule is nearly vibrationless, as can be arranged by providing a gas pressure which will give, during the lifetime of the excited molecule, sufficient inelastic collisions to remove the excess of vibrational energy communicated to the excited molecule in the excitation process. The purely electronic disturbance which occurs in the formation of the ordinary molecule from a vibrationally quiescent, electronically excited molecule can excite only a limited number of vibrations in the ordinary molecule, these being determined by the difference of shape between the excited molecule and the ordinary molecule. Thus we obtain the comparatively simple spectrum, shown in fig. 15, in which certain regularities are obvious even to casual inspection. On the other hand, if the electronically excited molecule is vibrating with the whole of the excess of energy communicated in the excitation process (as can be arranged by maintaining so low a gas pressure that there is no opportunity for inelastic collisions during the life of the excited molecule), then the regenerated ordinary molecule will possess many vibrations, generated in the vibratory motion of its precursor. We thus obtain a complicated spectrum, such as is shown in fig. 16. Such spectra are more difficult to analyse, but when analysed yield more information, than do spectra of the simpler type.

These spectra record a number of vibration frequencies, which we already knew from the Raman and infra-red spectra, but in addition they yield a few others. One is known to be the butterfly-like puckering vibration of the carbon ring (fig. 12, formula 15). Another is probably to be identified with the rotatory vibration of the hydrogen atoms in the plane of the ring (fig. 12, formula 3).

The second method of searching for the frequencies of the inactive vibrations is by examining the Raman and infra-red spectra of benzenes in which only some of the positions are filled by deuterium atoms. By judicious partial deuterium substitution, one can destroy chosen elements of molecular symmetry, and this allows certain previously forbidden lines and bands to appear. It is sometimes possible to identify the newly appearing frequencies by studying the types of substitution which allow them to appear; and, of course, confirmation can always be obtained from the frequency shifts provided the frequencies can be rendered active in more than one benzene. In illustration of the former method of identification

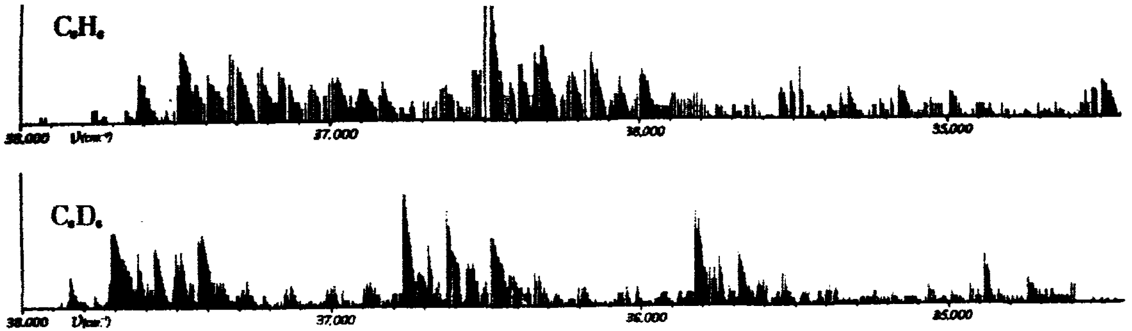


FIG. 15. Part of high-pressure fluorescence spectra of benzene and hexadeuterobenzene (vapour).

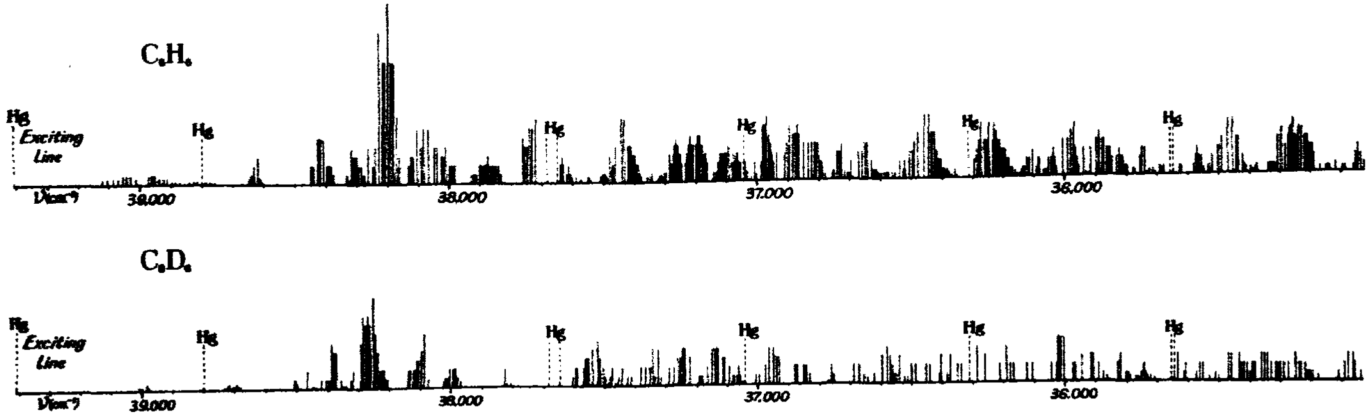
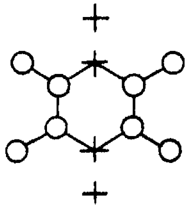
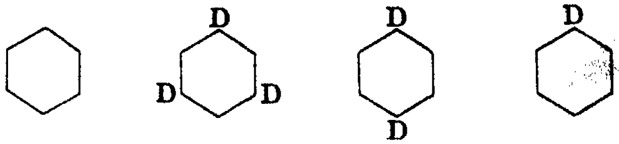


FIG. 16. Part of low-pressure fluorescence spectra of benzene and hexadeuterobenzene (vapour) excited by Hg 2537 A.

we may consider the "butterfly" vibration (fig. 12, formula 15). When we work out the selection rules which govern the appearance of this vibration in the Raman and infra-red spectra of a number of benzenes having different dispositions of deuterium atoms, for instance the four benzenes formulated below, we find that it should appear in some spectra and not in others (in one it should appear as a doublet), as is indicated under the formulae. No other vibration has these properties in their entirety, excepting one of high frequency which cannot be confused with the necessarily low frequency of a ring-puckering vibration. When we examine the spectra, we find that one, and only one, low frequency has these properties. Knowing the laws governing the isotopic frequency shifts, it can be deduced from the observations that in ordinary benzene this vibration has the frequency 398 cm.$^{-1}$.



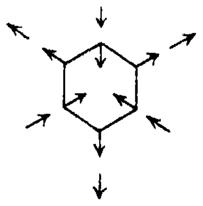


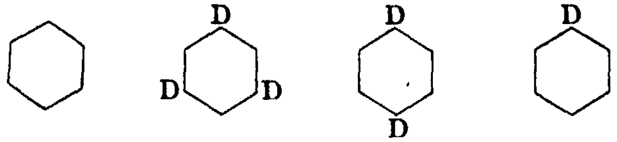
Theory				
Raman	Forbidden	Allowed	Forbidden	Allowed Double
Infra-red	Forbidden	Forbidden	Allowed	Allowed
Observed frequencies	—	375	383	382, 398

∴ Frequency in C_6H_6 = 398 cm.$^{-1}$.

Another vibration whose frequency is easily found in just the same way is the planar, trigonal, bending vibration of the carbon ring (fig. 12, formula 5). The selection rules for the same series of benzenes are shown below, together with the numerical observations. From these it can be deduced that in ordinary benzene this vibration has the frequency 1007 cm.$^{-1}$ (see opposite, p. 173).

The determination of the frequencies of the vibrations which are spectroscopically inactive in benzene is of importance, because only when the frequencies are known can their inactivity be conclusively established: it is necessary to be sure that they are not merely obscured under some other line or band. The two vibrations last mentioned have been selected for consideration, because their complete exclusion from the spectra of ordinary benzene combined with the absence of coincidences, affords a self-contained proof (which can be considerably supplemented) that the undisturbed benzene molecule possesses very strictly every one of the elements of symmetry of the plane, regular hexagonal model. The quantitative frequency shifts, as well as detailed measurement of the infra-red bands, give information concerning the dimensions (C—C, 1·39 A; C—H, 1·08 A), the carbon-carbon distance being already known from diffraction data.





Theory				
Raman	Forbidden	Allowed polarized	Forbidden	Allowed polarized
Infra-red	Forbidden	Forbidden	Allowed	Allowed
Observed frequencies	—	1001	993	1008

∴ Frequency in $C_6H_6 = 1007$ cm.$^{-1}$.

We do not yet know a sufficient number of the frequencies of the partly deuterated benzenes to undertake more than a rough calculation of the force field within the molecule.

Although the benzene problem is not yet fully solved, it seems in recent years to have emerged from a period of vague speculation and qualitative experiment, and to have entered on a period of much improved, if still imperfect, theory, supported by concordant and quantitative observation.

The photography of airscrew sound waves

By W. F. Hilton, Ph.D.

(*Communicated by L. Bairstow, F.R.S.—Received* 15 *January* 1938—
Revised 17 *August* 1938)

[Plates 3–8]

1. Introduction

The detail of the problem of airscrew noise has received little attention in the past, and the present research was undertaken to investigate pressure waves in the neighbourhood of an airscrew. It appears that airscrew noise may be considerably reduced by decreasing the tip speed, but the method involves an increase of airscrew diameter which conflicts with other aspects of efficient design.

The effects now considered are those due to the passage of an aerofoil through air, and the fundamental problem examined has wider application than the consideration of noise production. The primary observations are photographs of the pressure wave system caused by an airscrew blade in motion at high speeds; from the wave shape so found, and measurements of the width of the image a method is developed for estimating the pressure amplitude of the wave.

2. Experimental method

The photographs of airscrew pressure waves reproduced in figs. 6–11 (Plates 3–8) were taken by Dvorak's direct shadow method, using a spark giving exposures of about a millionth of a second. This method has for many years been used for photographing bullets in motion. Nothing was added to the air to render the motion visible, and the airflow near the airscrew was in no way disturbed by the experimental technique. The images obtained by this method are free from distortion, however large the angle of field included, but they are magnified in the ratio

$$m = (f+b)/b, \tag{1}$$

where b is the distance from spark to airscrew, and f the distance from airscrew to plate. Thus the wave shape is faithfully recorded on the photograph if f is constant. The only optical defect of the image is a slight diffraction effect at the edges of all shadows.

The variable pitch airscrew used was two feet in diameter, driven by a compressed air turbine, and is described by Douglas and Hartshorn (1931). Tip speeds up to 1·2a, where a is the velocity of sound in undisturbed air, were obtained. The lowest tip speed at which sound waves were photographed was 0·83a, but eddies could be recorded at tip speeds as low as 0·4a. The airscrew operated at zero rate of advance through the air and was enclosed in a dark room. A secondary spark gap closed by the airscrew itself was used to ensure that the main spark passed only when the airscrew was in the desired position. Conditions did not repeat with sufficient accuracy to enable the images from successive sparks to be superimposed. An induction coil with a manually operated make and break produced the spark. When the spark gap was suitably adjusted, no spark was obtained at "make", but only at "break". The charge was held by a condenser until the airscrew came round into the correct position and produced a discharge in the main spark gap. A tell-tale spark gap in series with the others was placed outside the dark room, to indicate when a photograph had been taken.

3. DESCRIPTION OF PLATES

Typical photographs are reproduced in figs. 6–11 (Plates 3–8) showing the black shadow of the airscrew blade, together with the curved images of the sound waves. The photographs were taken from the front of the airscrew, and are mounted to correspond with the airscrew blade moving upwards. On many of the photographs the eddy sheet behind the airscrew blade is visible, as well as the trailing vortex from the airscrew tip. It is believed that these are the first direct photographs taken of vortex motion in air. At low speeds only the trailing vortex and eddy sheet are visible, but at a definite critical tip speed, a series of diffuse wavelets appears at the blade tip. At a slightly higher tip speed these wavelets condense into a single wave front, similar to that shown in fig. 7, Plate 4, taken at a speed $V = 0{\cdot}87a$. Fig. 8, Plate 5, was taken at a speed $V = 0{\cdot}984a$, and shows a well-developed wave. It also shows a wave emitted by a previous blade, which after reflexion at the wall of the room has returned in time to be recorded on the photograph. Measurements of this reflected wave enabled the wave velocity to be determined. The eddying wake and tip vortex are also clearly defined in fig. 8.

Fig. 9 (Plate 6), unlike the other photographs shown, was taken in the plane of rotation of the airscrew, the direction being inclined at about 30° to the length of the blade, giving a very foreshortened view. The extent of the wave in the fore and aft direction can be seen from such photographs, as well as the eddy sheet and tip vortex from the other blade. Figs. 9 and 10, Plates 6 and 7, were taken simultaneously at a speed $V = 1{\cdot}083a$ for the purpose of applying the method of pressure amplitude determination described in § 5. Fig. 10 shows a wealth of detail of reflected waves, some of which have been reflected several times. In front of the blade may be seen the faint eddying motion due to the passage of the previous blade. A bow wave is seen to be forming ahead of the blade, and is in two parts, which unite at a slightly higher speed. The speed at which a bow wave first appears is discussed in § 7. The black pointer indicates the position of the blade tip when the blade was horizontal and the shadow of a plumb-line is also included.

Fig. 11, Plate 8, was taken at the highest speed reached, $V = 1{\cdot}21a$, and shows a well-developed bow wave, also the "second" and "third" sections of the trailing wave, defined in fig. 2. The tip vortex due to the passage of the blade during the previous revolution, as well as that due to the other blade, are seen quite clearly on this plate. It is also instructive to note the abrupt change in the eddying wake, from a high-speed fine-grained, to a low-speed coarse-grained appearance. This change occurs at a radius corresponding to a linear speed $V = 0{\cdot}98a$, and is the only discontinuity observed at or near the speed of sound.

The plates were chosen as representative from over 200 photographs taken under various conditions. The blade section of the airscrew in figs. 7, 8 and 11 was R.A.F. 28 at 5° incidence at the tip, while fig. 6 was taken at 11° 40′ incidence. Figs. 9 and 10 were both R.A.F. 32 section at 10° 30′ tip incidence.

4. CALCULATION OF THE WAVE SHAPE

The curvature of these waves is due to the rotation of the airscrew, and was predicted by Lamb (1921).

If it be assumed that the wave was generated by a point source of sound at the tip of a blade of radius A, the wave front may be drawn by the simple geometrical construction shown in fig. 1. This construction assumes that when the airscrew was in a position making an angle θ with its present position, the tip emitted a spherical wave which has travelled a distance $A\theta/K$ before the photograph was taken, K being the ratio of the tip speed of the airscrew to the velocity of sound. The wave front will be the envelope of a series of circles, drawn for various values of θ. This construction assumes without proof that the wave is produced by a single source of sound at the airscrew tip, and it gives no envelope for tip speeds less than the speed of sound, although photographs such as fig. 7 show a very definite wave front at speeds as low as $0{\cdot}87a$.

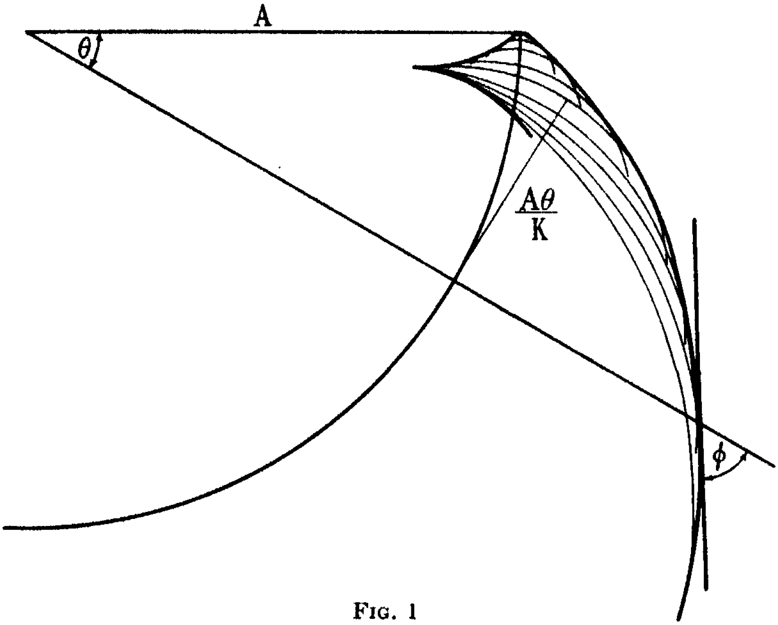


FIG. 1

Professor G. I. Taylor, F.R.S., in correspondence with the author, has given a mathematical expression for this wave shape by assuming

(i) That the wave is propagated with velocity (a).

(ii) That the wave pattern rotates with angular velocity ω about the hub. Referring to fig. 1, ϕ is the angle between the wave front and the radius vector r, which makes an angle θ with an arbitrary fixed direction. Then from assumption (ii),

$$\cos\phi = \frac{a}{r\omega}, \tag{2}$$

where (a), the velocity of the sound wave becomes equal to a_0 for small amplitudes. Remembering that $\tan\phi = \frac{r\,d\theta}{dr}$, elimination of r and integration of the resulting relation between θ and ϕ, gives

$$\theta + \text{constant} = \tan\phi - \phi. \tag{3}$$

The shape of the wave may now be plotted, by taking a series of arbitrary values for ϕ, finding r from (2), and θ from (3).

Professor Taylor ended his mathematical treatment of the wave shape by remarking that it gave the same result as the geometrical construction outlined above, *without* assuming a point source of sound at the blade tip. However, if a curve be plotted from (2) and (3), taking the constant as zero, a right and left handed pair of spirals starting from a cusp are obtained, as is shown in fig. 2. To reproduce Lamb's geometrical construction, it must further be assumed that the wave comes from the blade tip and two alternative positions of the cusp must be combined, treating the part of the wave ahead of the blade as imaginary. This imaginary part arises from the mathematical method, and is shown dotted in fig. 2, together with a convenient nomenclature for the various wave fronts. The function of the blade tip in the production of these waves would seem to be doubtful, and the assumption made to reproduce Lamb's geometrical construction is quite arbitrary. It is better to regard these spiral curves as the only wave pattern which can rotate with given angular velocity about the hub without change of shape, and to leave the relative positions of the airscrew blade and wave pattern to be determined by experiment.

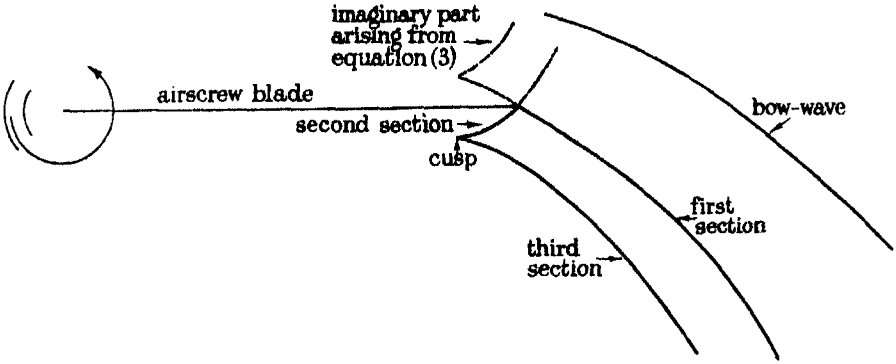


FIG. 2. Nomenclature adopted for distinguishing various wave fronts.

The cusp from which the two spirals spring is at such a radius that it moves round the hub with exactly the velocity of sound and thus for lower speeds of rotation, the cusp is farther from the centre. When the tip speed falls below the speed of sound, the cusp and the rotating wave system will be entirely outside the airscrew disk, and the wave cannot join the airscrew at any point. Since it is known that conditions favourable to the formation of a shock wave exist on the blade alone at subsonic velocities, it is evident that some additional assumption must be made to explain the wave shapes seen in figs. 7 and 8 at subsonic velocities. Since the two spirals meeting in a cusp are the only waves which can be propagated without change of shape, it is assumed that the cusp prolongs itself and joins on to the airscrew blade, and that the shape of this connecting wave is continually altering. The upper spiral or "second section" starting from the cusp in fig. 2 does not exist at subsonic tip speeds, leaving the lower spiral, or "third section" to be joined to the airscrew by this undetermined extension of the wave. Fig. 7, taken at a speed $0{\cdot}87a$, shows this connecting part of the wave moving faster than the blade, and getting ahead of it.

On comparing the theoretical curves with the actual wave shapes, the best agreement is found by comparing photographs taken at one speed with theoretical curves for a rather higher speed, as may be seen from the tracings to fit figs. 10 and 11, where the higher speed curve is shown dotted. This effect is almost certainly due to the fact that the air in the neighbourhood follows the blade, and so augments the speed of propagation of the waves relative to still air.

An estimate of the velocity of the air may be obtained by finding the value of the speed constant a/ω which must be substituted in equation (2) to obtain the best fit between theory and experiment. Since the rotational speed ω is fixed, the magnitude of a, the enhanced velocity of propagation is found. Thus for fig. 11, taken at a tip speed of $1{\cdot}21a_0$, the dotted curve calculated for a speed $1{\cdot}3a_0$ gives the best fit. (The dotted curve for $V = 1{\cdot}1a$, fitting the bow wave, is discussed later.) Hence $a = 1{\cdot}074a_0$ and the velocity component of the air perpendicular to the wave front is $0{\cdot}074a_0$ or 81 ft./sec. The component of air velocity along the wave front is not determined by this method, but should be fairly small. An air speed of 62 ft./sec. was obtained similarly for fig. 8, which was taken at the same blade incidence as fig. 11, but at the lower tip speed of $0{\cdot}984a_0$. Figs. 7 and 10 were taken at greater blade incidences (see § 3), and gave higher air speeds of 101 and 129 ft./sec. respectively. These results are in accordance with the velocities one would expect from general knowledge of the flow near an airscrew.

It may seem surprising that the wave due to any one blade can be affected by the disturbances due to the same blade, but this is explained by the fact that the "second section" is propagated inwards towards the boss, and thus finds itself in the eddying wake. The third section, starting from the cusp, is thus also immersed in the moving wake, until it has propagated for a sufficiently long time to enable it to get clear of the wake.

Since the first section is propagated in undisturbed air, it would be expected to have the normal velocity, and to fit the theoretical curves exactly. An excellent fit is obtained for fig. 10, the only plate reproduced which shows the first section at all distinctly. The reason for the absence or faintness of the first section in all cases may be that the blade tip alone is instrumental in producing it, since it only arises as a consequence of assuming that the wave front comes from the blade tip.

The bow wave would also be expected to fit the theoretical curves. Such agreement is found in fig. 10, where the bow wave is newly formed, and of small amplitude. In fig. 11 the bow wave is much stronger, and it is found to give a better fit with a theoretical curve drawn for the lower speed $V = 1{\cdot}1a$. This apparent discrepancy may be explained by the effect of finite amplitude on the velocity of wave propagation. If the wave had an amplitude of 0·89 atm., its velocity from equation (12) would be $a = 1{\cdot}1a_0$, and the curve for $V = 1{\cdot}1a$ would give the correct shape for a speed $V = 1{\cdot}21a_0$. It thus appears that the wave shapes observed can be fairly well accounted for on a theoretical basis.

5. DETERMINATION OF THE PRESSURE AMPLITUDE OF A WAVE, FROM MEASUREMENTS OF A PHOTOGRAPH

Consider a surface of revolution of radius r, separating compressed air of refractive index $n + \delta n$ from air of refractive index n. S is a point source of light, and CD a photographic plate. All rays above SAD in fig. 3 will fall on the plate without deviation, to produce a uniform grey fog. Rays falling below SAD will be deviated away from D, as is also shown in fig. 3. There will thus be a shadow of width d below D, limited at a point C by the ray of light having minimum deviation. The author is indebted to Professor

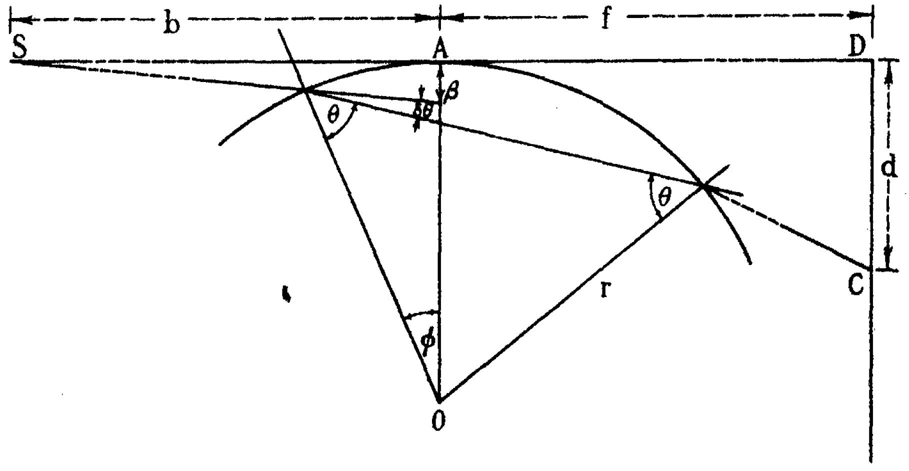


FIG. 3

Fig. 6. $\Gamma/a = 0{\cdot}83$.

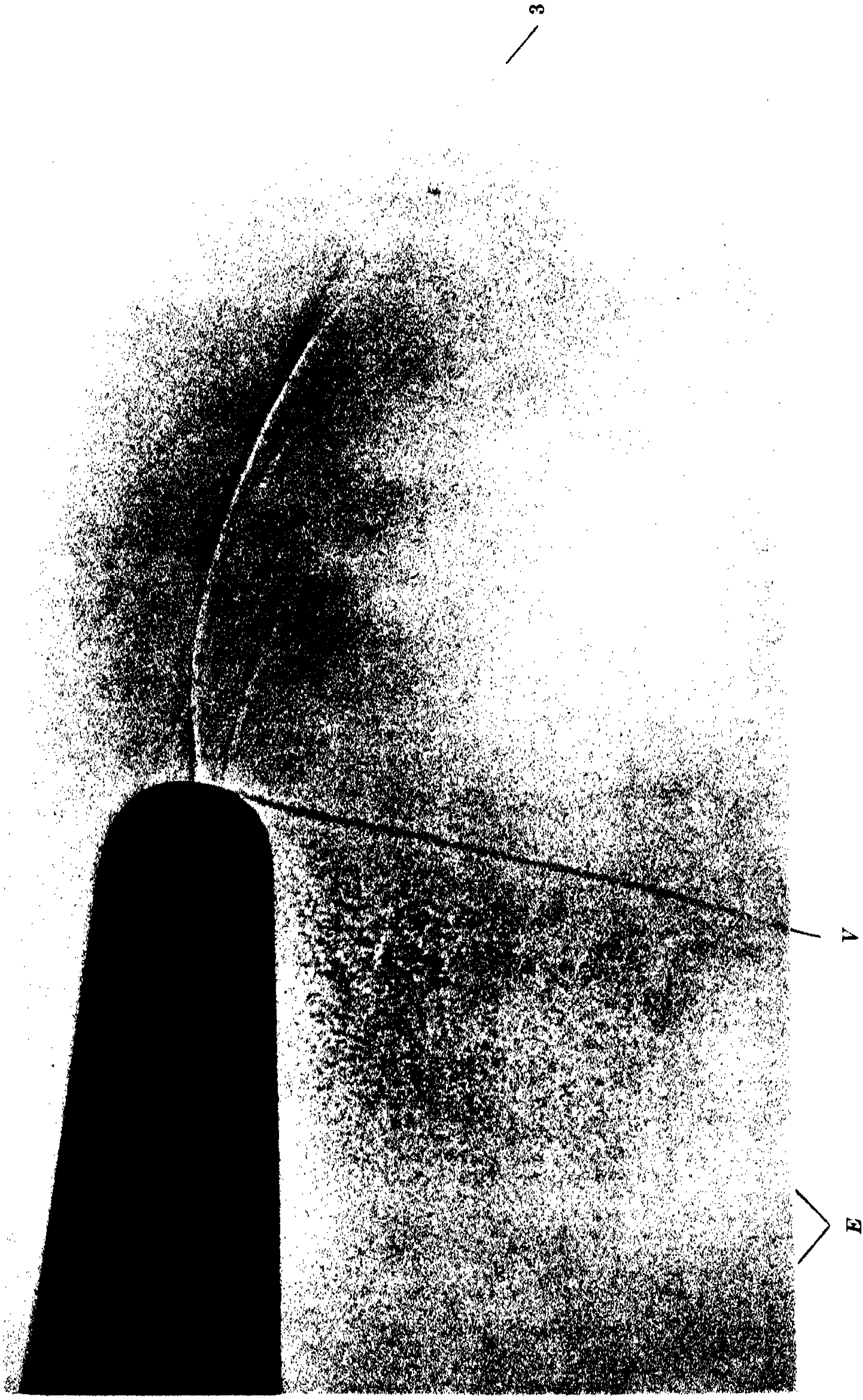


FIG. 7. $V/a = 0{\cdot}87$.

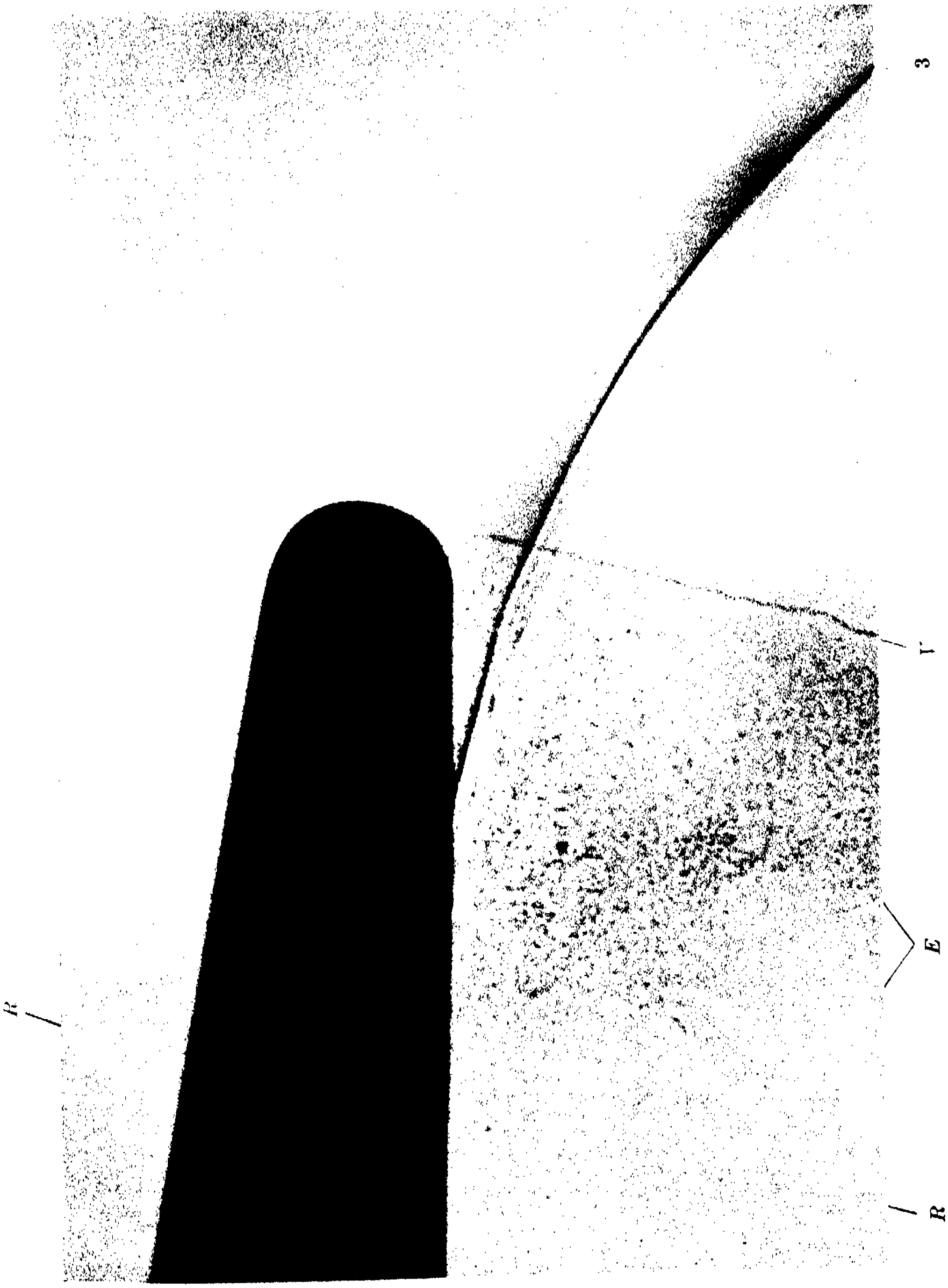


FIG. 8. V/a = 0·984.

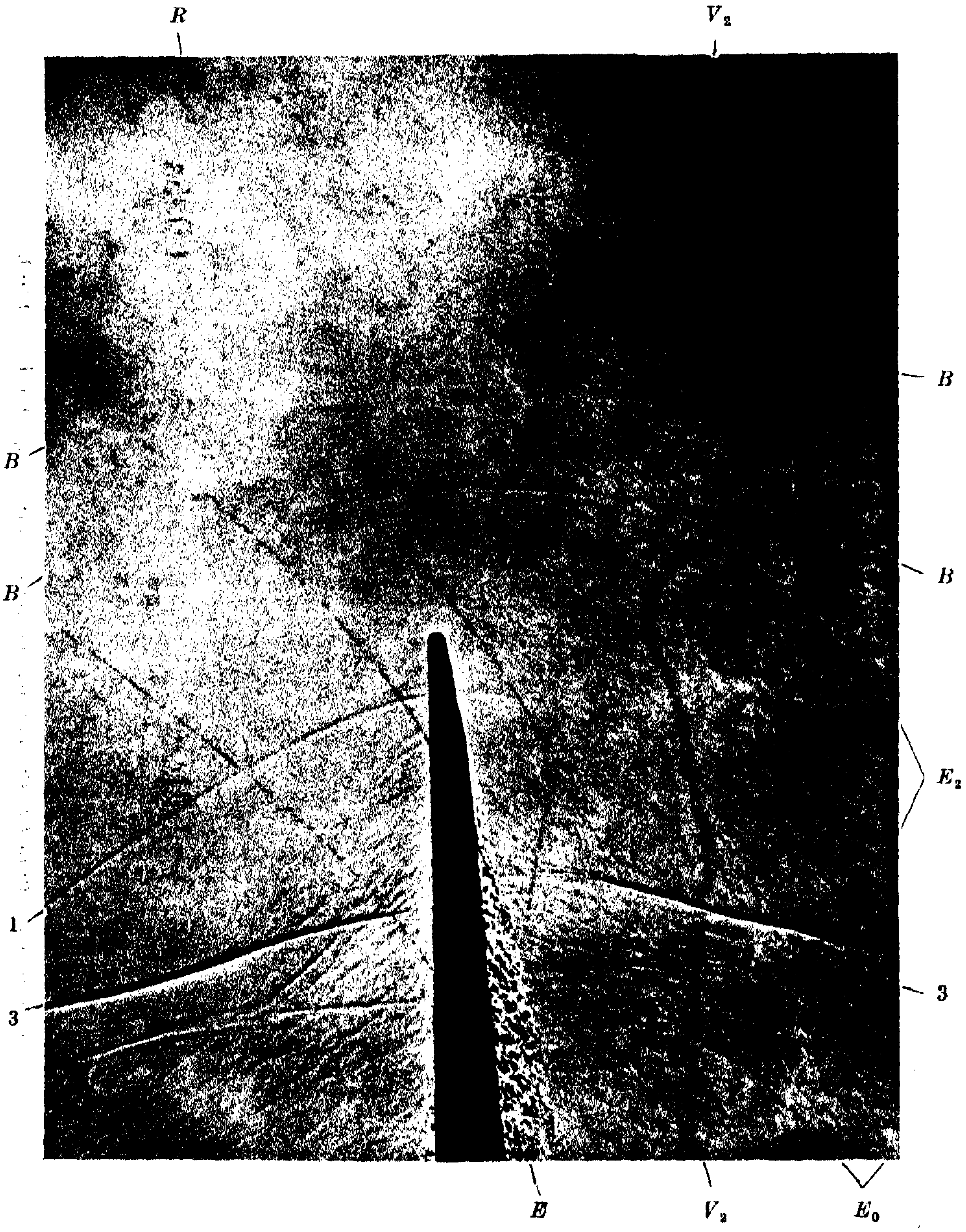
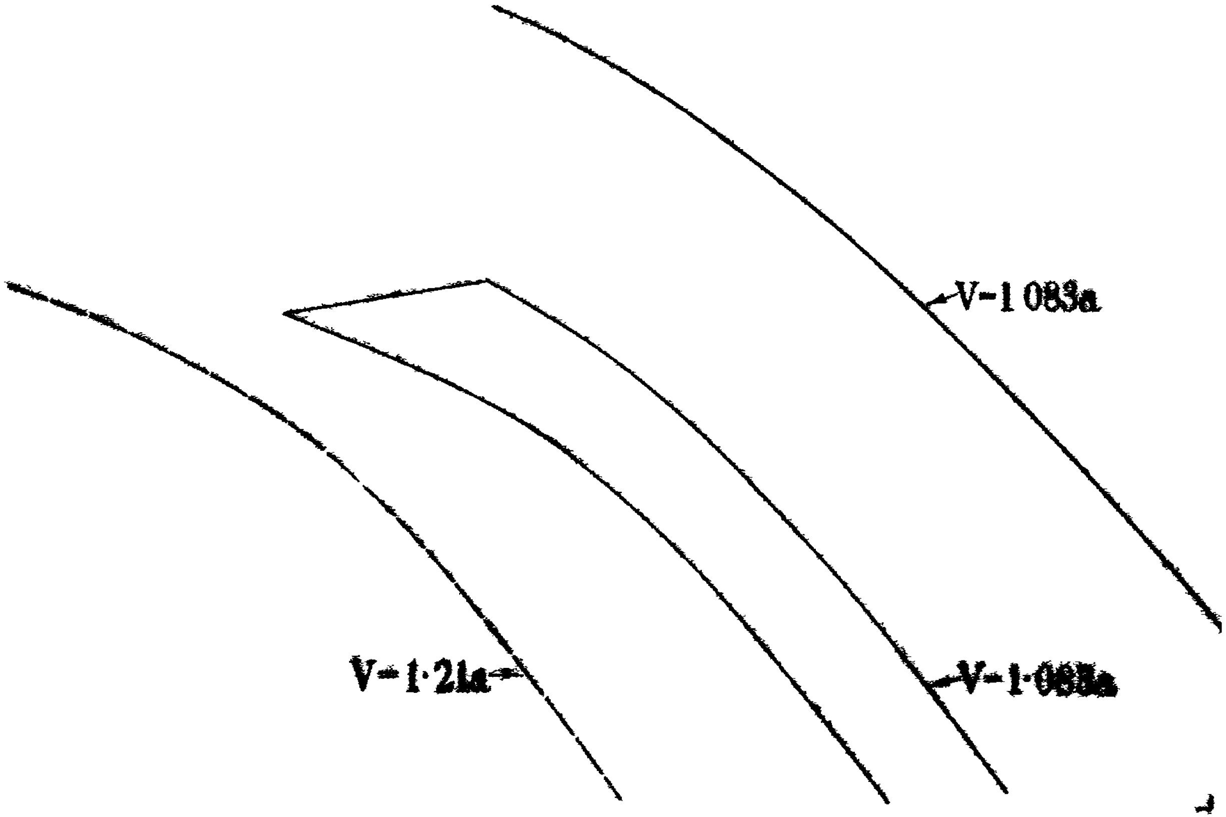


FIG. 9. $V/a = 1{\cdot}083$.



Transparency to fit fig. 10, showing theoretical wave shape for $V/a = 1{\cdot}083$.

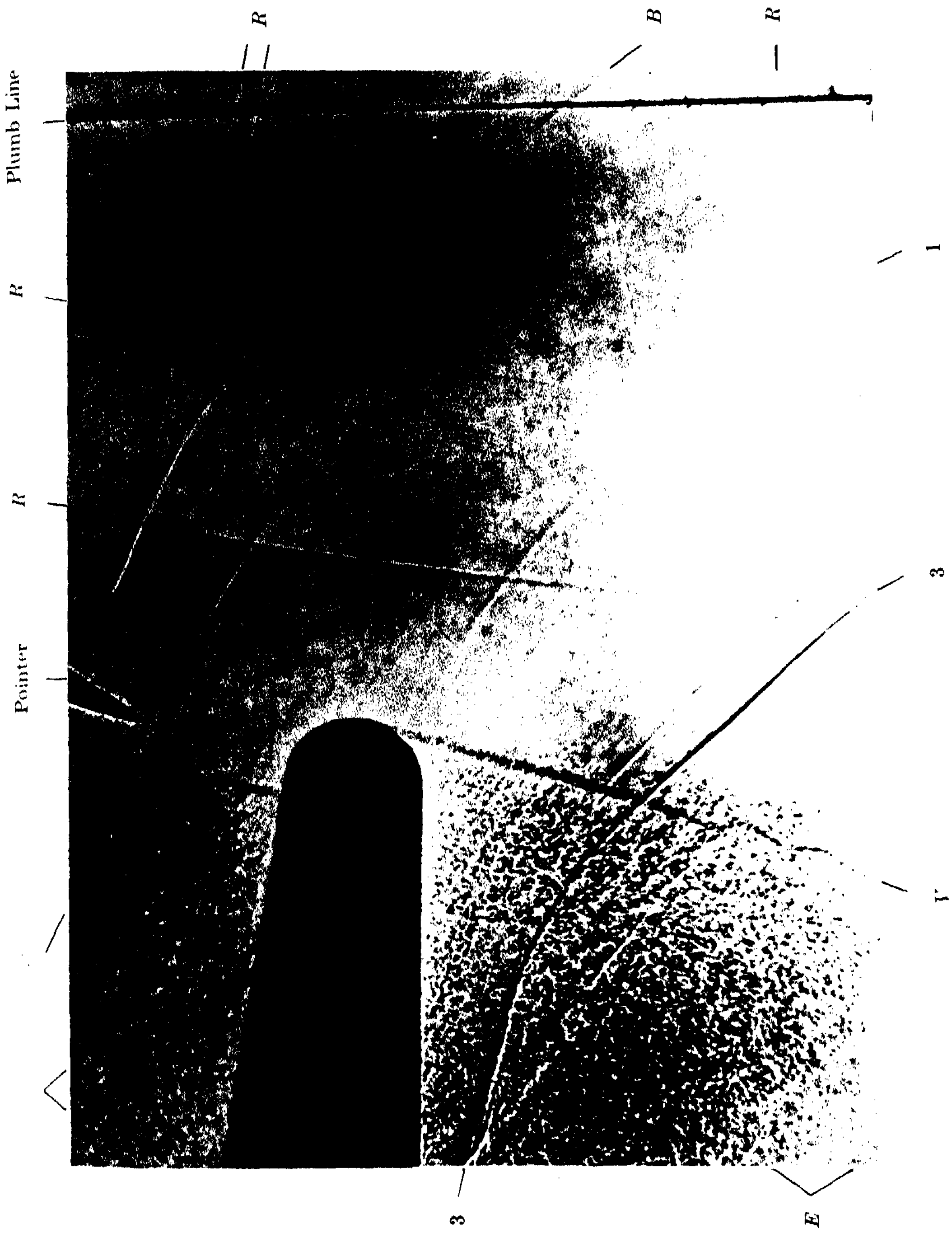


FIG. 10. $V/a = 1{\cdot}083$.

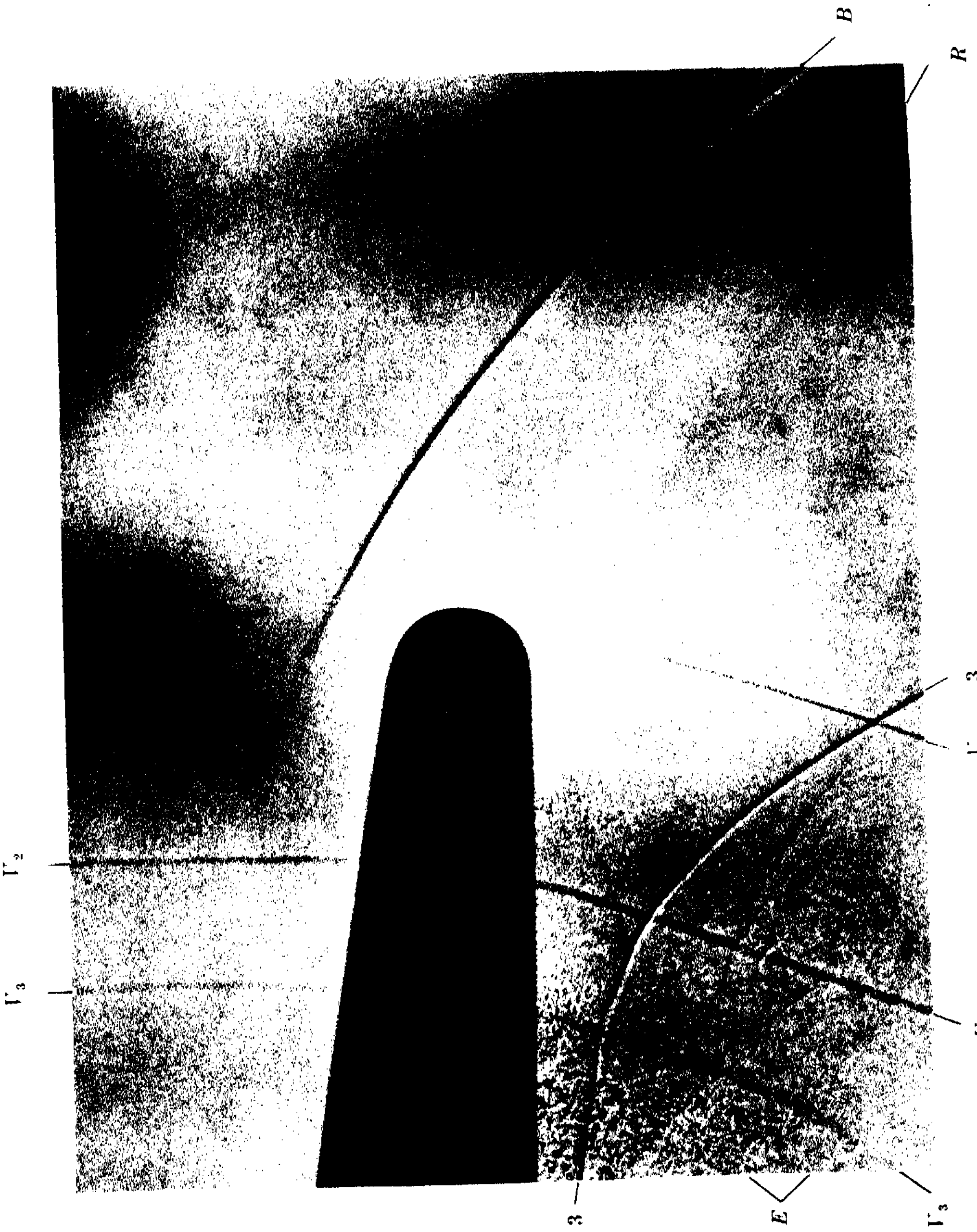
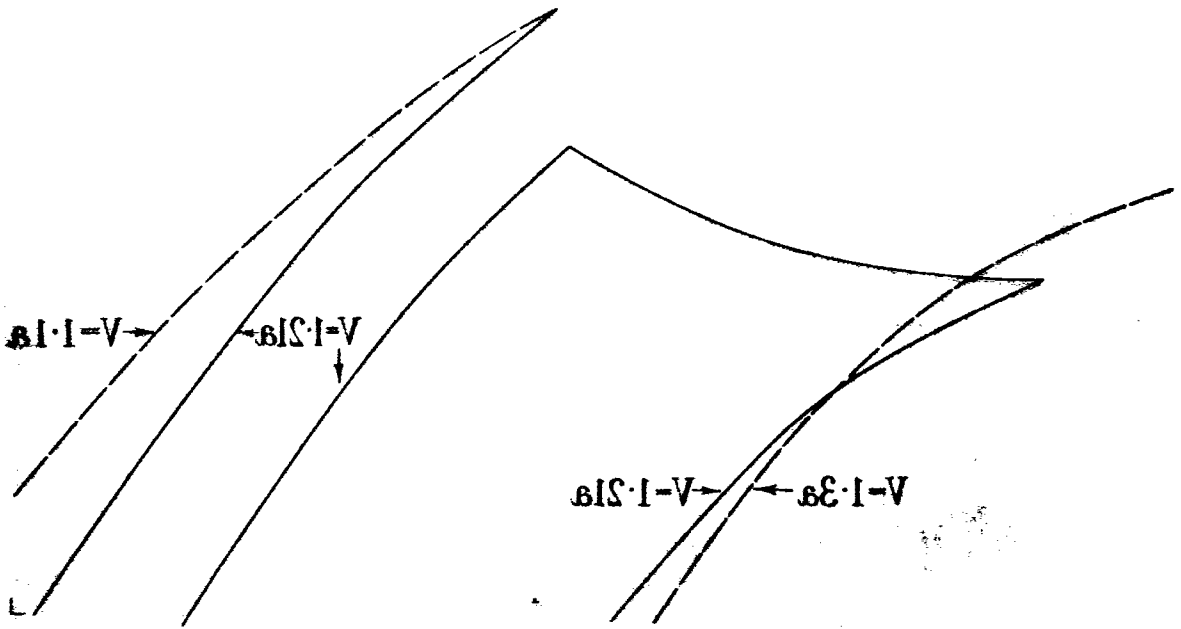


Fig. 11. $V/a = 1{\cdot}21$.



Transparency to fit fig. 11, showing theoretical wave-shape for V/a = 1·21.

G. I. Taylor, F.R.S., for pointing out that this ray of minimum deviation is not necessarily the ray having grazing incidence. It should be noted that fig. 3 is greatly exaggerated, for the sake of clearness. ϕ, β, $\delta\theta$, and δn are all small quantities, but of different orders and it is necessary to be careful when neglecting any of them. Thus ϕ is of square root order in comparison with β. Using Snell's law at the first refraction, and remembering

$$\phi+\theta+\delta\theta+\beta/b = 90^\circ,$$

we have
$$\frac{n}{n+\delta n} = \frac{\sin\theta}{\sin(\theta+\delta\theta)} = \frac{\cos(\phi+\beta/b+\delta\theta)}{\cos(\phi+\beta/b)},$$

or
$$1-\frac{\delta n}{n+\delta n} = \cos\delta\theta-\sin\delta\theta\tan(\phi+\beta/b).$$

Since δn and $(n-1)$ are small, it is permissible to write δn for $\delta n/(n+\delta n)$, and to expand $\cos\delta\theta$ and $\sin\delta\theta$, giving

$$2\delta n = \delta\theta^2+2\delta\theta\tan(\phi+\beta/b). \tag{4}$$

Further, from the geometry of the circle,

$$2r\left(\beta-\frac{\beta r}{b}\tan\phi\right) = r^2\tan^2\phi,$$

hence
$$\tan\phi = -\frac{\beta}{b}\pm\sqrt{\left(\frac{\beta^2}{b^2}+\frac{2\beta}{r}\right)}.$$

Since β/b is small, although ϕ is not so small, write $\tan(\phi+\beta/b) = \tan\phi+\beta/b$. and neglect β^2 in comparison with β. The negative sign of the root applies only to the case of a second refraction, and thus

$$\tan(\phi+\beta/b) = \sqrt{\frac{2\beta}{r}}. \tag{5}$$

Substituting in equation (4) and neglecting $\delta\theta^2$,

$$\beta = \frac{r}{2}\left(\frac{\delta n}{\delta\theta}\right)^2-\frac{r\,\delta n}{2}. \tag{6}$$

The deviation d of a ray of light is given by

$$d = m\beta+2f\delta\theta, \tag{7}$$

where m is the optical magnification. Substituting for β from (6),

$$d = \frac{mr}{2}\left(\frac{\delta n}{\delta\theta}\right)^2+2f\delta\theta-\frac{mr\,\delta n}{2}. \tag{8}$$

For minimum deviation

$$\frac{d(d)}{d(\delta\theta)} = -\frac{mr\,\delta n^2}{\delta\theta^3} + 2f = 0,$$

i.e.
$$\frac{\delta n^2}{\delta\theta^3} = \frac{2f}{mr}, \quad \text{or} \quad \frac{\delta n^2}{\delta\theta^2} = \frac{2f\,\delta\theta}{mr}.$$

Substituting these values in (8), and remembering that d is now the special value for minimum deviation,

$$d = f\,\delta\theta + 2f\,\delta\theta - \frac{mr\,\delta n}{2} \quad \text{or} \quad 2d + mr\,\delta n = 6f\,\delta\theta.$$

Cubing both sides of this equation, and substituting for $\delta\theta^3$

$$8d^3 + 12d^2 mr\,\delta n + 6dm^2r^2\,\delta n^2 + m^3r^3\,\delta n^3 = 108f^2\,\delta n^2 mr.$$

In experiments, the following approximate values were found:

$$d = 0{\cdot}07, \quad m = 1{\cdot}5, \quad r = 20, \quad \delta n = 2 \times 10^{-5}, \quad f = 50.$$

Thus all terms but the first and last may be neglected, giving

$$\delta n = \frac{1}{f}\sqrt{\frac{2d^3}{27mr}}. \tag{9}$$

In order to relate δn to the pressure amplitude δp of a sound wave, we have $(n-1)/\rho$ = constant, and the Rankine-Hugoniot relation connecting pressure and density in a shock wave. These may be written

$$(n_0 - 1)\left(\frac{1}{\rho_0} - \frac{1}{\rho}\right) = \frac{\delta n}{\rho}$$

and
$$\left(\frac{1}{\rho_0} - \frac{1}{\rho}\right)\left(\frac{2\gamma p_0}{\gamma - 1} + \delta p\right) = \frac{2}{\gamma - 1}\frac{\delta p}{\rho} \text{ respectively,}$$

where δp is the pressure amplitude of the shock wave. Eliminating ρ_0 between these equations,

$$\delta n[(2\gamma p_0/\delta p) + (\gamma - 1)] = 2(n_0 - 1). \tag{10}$$

Eliminating δn between (9) and (10),

$$\frac{p_0}{\delta p} = \frac{f(n_0 - 1)}{2\gamma}\sqrt{\frac{54mr}{d^3}} - \frac{\gamma - 1}{2\gamma}, \tag{11}$$

giving the pressure amplitude δp of a shock wave in terms of experimentally known quantities.

To measure the width d and the radius of curvature r of a shock wave, it was necessary to take a pair of photographs simultaneously in two mutually perpendicular directions, as in figs. 9 and 10. Measurement of f and of b gives the optical magnification m, by using equation (1), and, taking $\gamma = 1\cdot402$ and $n = 1\cdot000303$,* the pressure amplitude may be calculated from (11). A series of such pressure amplitude calculations is given in Table I. On photographs where more than one wave was visible, the final column gives the total amplitude for all waves from the blade.

TABLE I. CALCULATION OF THE PRESSURE AMPLITUDE OF THE WAVES (δp)

Rev./ min.	b in in.	r in in.	d in mm.	f in in.	$m = 1 + (f/b)$	$\delta p/p_0$	Total	Aerofoil section
9,900	40·3	6·7	0·35	16·4	1·406	0·0403	0·0403	R.A.F. 32, 1·5°
10,800	39·6	8·6	0·55	17·1	1·432	0·0665	0·1115	
	39·3	8·6	0·43	17·4	1·443	0·0450		
11,700	38·3	9·0	0·68	18·4	1·481	0·0810	0·0810	
12,600	37·8	7·0	0·75	18·9	1·500	0·1030	0·1956	
	37·7	10·0	0·79	19·0	1·503	0·0926		
9,900	40·7	9·0	0·46	16·0	1·392	0·0538	0·0538	R.A.F. 32, 10·5°
10,800	39·0	6·4	0·46	17·7	1·453	0·0564	0·0564	
11,700	37·8	3·5	0·41	18·9	1·500	0·0593	0·1871	
	38·7	6·0	0·56	18·0	1·465	0·0764		
	38·5	20·0	0·39	18·2	1·472	0·0242		
	38·4	7·0	0·32	18·3	1·476	0·0312		
12,600	38·0	4·0	1·05	18·7	1·491	0·2241	0·2241	

This method of estimating pressure amplitudes of shock waves may be checked by using it on a bullet shock wave of known amplitude. The Ordnance Committee, Woolwich Arsenal, kindly lent the author a photograph of such a wave. The known velocity of the bullet being V, the value of a/V was found from measurements of the photograph, giving (a), the wave velocity. By using the Rankine-Hugoniot relation for shock waves, the wave velocity may be correlated with the pressure amplitude $(p - p_0)$, and for small amplitudes, it can be shown that

$$\frac{a}{a_0} = 1 + \frac{p - p_0}{7p}. \tag{12}$$

---

* A special experiment was made, to determine the refractive index of air appropriate to these spark photographs. A spectrogram was obtained of the light used for taking the photographs, on the same brand of plates as were used in the rest of the work. An analysis of this spectrogram gave a mean wave-length $\lambda = 3200$. The refractive index of air for this wave-length is 1·000303, and this value was employed in the calculations.

Measurement of the photograph gave $a/a_0 = 1{\cdot}087$, whilst the value of a/a_0 found by the present method was 1·041.

The conventional method depends on the determination of a small difference between the Mach angle ($\sin^{-1} a/V$) and the measured angle ($\sin^{-1} a_0/V$), while the present method depends on the measurement of the width of the dark band in the shock wave, which is not very distinct. It is probable that the discrepancy is due to inaccuracy in the latter measurement. The present method leads to a direct determination of the pressure amplitude, and would seem to be better for small amplitude waves, when the Mach angle becomes nearly equal to the angle of the wave. In the case of irregular shock waves, or shock waves propagated in moving air, as in these experiments, the conventional method breaks down entirely.

It was hoped to convert the pressure amplitudes shown in Table I into loudness of sound in decibels, when large errors in pressure amplitude would lead to small errors of loudness, but no method has yet been found for estimating the angle of the cone through which an airscrew blade emits sound at any given instant. The angle in a direction parallel to the axis of rotation is readily measured, but that in the plane of rotation of the airscrew is as yet indeterminate. Experiment would seem to show that it is very small, about one-fortieth of the spreading in the other direction.

The loudness of the sound was measured with a Barkhausen audiometer. The results are plotted in fig. 5, and are discussed in § 8 below.

6. DETERMINATION OF THE VELOCITY OF PROPAGATION

It is desirable to know the velocity of propagation, for as is shown in § 5, these waves possess a small but finite amplitude, which might be sufficient to make this velocity appreciably greater than normal. In addition, the velocity usually quoted is for pure air, but in the experiments a certain amount of oil spray was inevitably present in the air at these high speeds. Incidentally, this oil spray caused much trouble by spotting the plates.

On certain of the negatives waves reflected from the walls of the room are to be seen. Measurement of their very slight curvature enabled the distance and direction of the reflecting surface to be estimated. A plane board was then nailed over the suspected surface, and the photograph repeated. A displacement of the reflected wave by a distance equal to twice the optical magnification times the thickness of the board indicated that the true reflector had been found. The distance of the board from the airscrew was then accurately measured. In order to determine the velocity, the angle through which the airscrew had turned since emitting the wave must be determined, as well as the distance travelled by the wave after emission.

The particular plate from which the wave velocity was calculated has not been reproduced, but fig. 8 was taken under similar conditions. Fig. 4 shows diagrammatically the reflected wave, which intersects the axis OG of the airscrew blade at a point A. This wave has been reflected at an angle δ from the wall at B. The original, unreflected sound wave SSS, is assumed to retain its position and shape relative to the airscrew blade, as the latter rotates. Draw a tangent to the wave front, from the tip of the blade and consider a position OD of the airscrew, such that this tangent DE to the wave front $S'S'S'$ is perpendicular to EB, the line of propagation of the wave before reflexion.

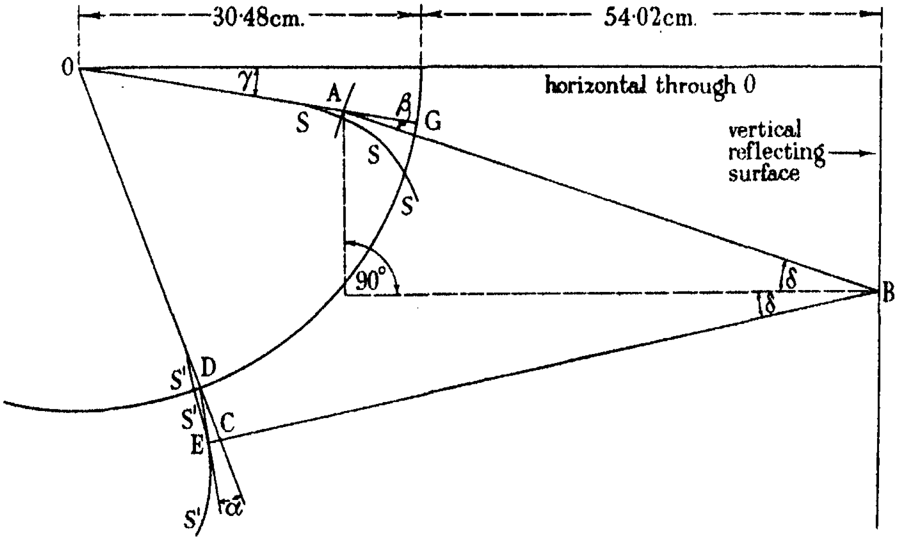


FIG. 4

The wave will move along ECB, and after reflexion at B will travel along BA, to arrive at A in time to be photographed. During this time the airscrew has not necessarily turned through an angle θ, but possibly through $(180n+\theta)^\circ$, where n is any positive integer. From consideration of the dimensions of the apparatus, it was evident that n was unity. From the geometry of the figure,

$$\theta = 90^\circ - \beta - 2\gamma - \alpha, \tag{13}$$

and hence from measurements of the photograph it was found that $\theta = 31{\cdot}6^\circ$. Since $n = 1$, the airscrew had turned through 211·6° whilst the wave had travelled a distance $(EB + BA)$, which was calculated to be 116·5 cm.

These quantities lead to a wave velocity $a = 352$ m./sec. Since the air temperature was 29·2° C, the normal velocity of sound was 349·1 m./sec., and apparently the finite amplitude of the wave caused a 1 % increase of velocity. If $a/a_0 = 1{\cdot}01$, from (12) $\delta p/p_0 = 0{\cdot}07$, which agrees reasonably well with Table I. A small correction should strictly be applied to the velocity, to allow for the effect of the wake through which the wave had to pass for a short distance, but this has not been determined.

7. Speed at which a bow wave first appears

According to the conventional theory of shock waves, a bow wave of infinitesimal amplitude should be formed at an infinite distance ahead of the blade as soon as the velocity of the airscrew tip equals that of sound, coming progressively nearer to the blade and increasing its amplitude as the speed is raised.

It is evident from the photographs taken, that a bow wave is not formed until the speed of sound at the tip is appreciably exceeded, the exact speeds being given in Table II. The bow wave has a broken and diffuse appearance when first formed, as may be seen from fig. 10, where it appears as two separate wavelets.

Table II

Geometrical incidence of blade tip	− 3° 35′	5° 55′	12° 35′	1·5°	10·5°
Blade section used	R.A.F. 28	R.A.F. 28	R.A.F. 28	R.A.F. 32	R.A.F. 32
Critical speed (series of diffuse wavelets first visible)	$0{\cdot}878a$	$0{\cdot}851a$	$0{\cdot}830a$	$0{\cdot}860a$	$0{\cdot}832a$
Nature of vortex	Just visible at $1{\cdot}04a$	Strong at $1{\cdot}0a$. Distinct at $0{\cdot}7a$	Very strong at $1{\cdot}0a$. Distinct at $0{\cdot}5a$	Just visible at $1{\cdot}0a$	Strong at $0{\cdot}83a$
"Bow wave" first appears faintly	$1{\cdot}11a$	$1{\cdot}08a$	$1{\cdot}11a$	Between $1{\cdot}0a$ and $1{\cdot}08a$	$1{\cdot}08a$

The experimental fact that a bow wave was not formed until the speed of sound was appreciably exceeded is probably explained by the peculiar flow in the neighbourhood of the tip of a high-speed airscrew working at zero advance. Exploration of the flow with silk streamers showed that the air was being sucked in at both sides of the airscrew disk near the tip, and ejected radially. For the air to be flung out radially, it must first have acquired considerable circumferential velocity, which would diminish the speed of the blade relative to the air through which it was passing, and delay the formation of a bow wave. If the effective incidence of the tip were increased, this induced circumferential velocity would be greater, and the formation of a bow wave would be further delayed. The data given in Table II will be seen to support this argument, when it is remembered that the geometrical incidence is a rough guide to effective incidence.

8. APPLICATION TO THE PROBLEM OF AIRSCREW NOISE

A remarkable feature of the audiometer readings is their continuity as the tip speed approaches and passes the speed of sound, as seen in fig. 5. Apparently the straight-line law connecting loudness in decibels with tip speed, mentioned by Capon (1933), may be extrapolated to apply to speeds

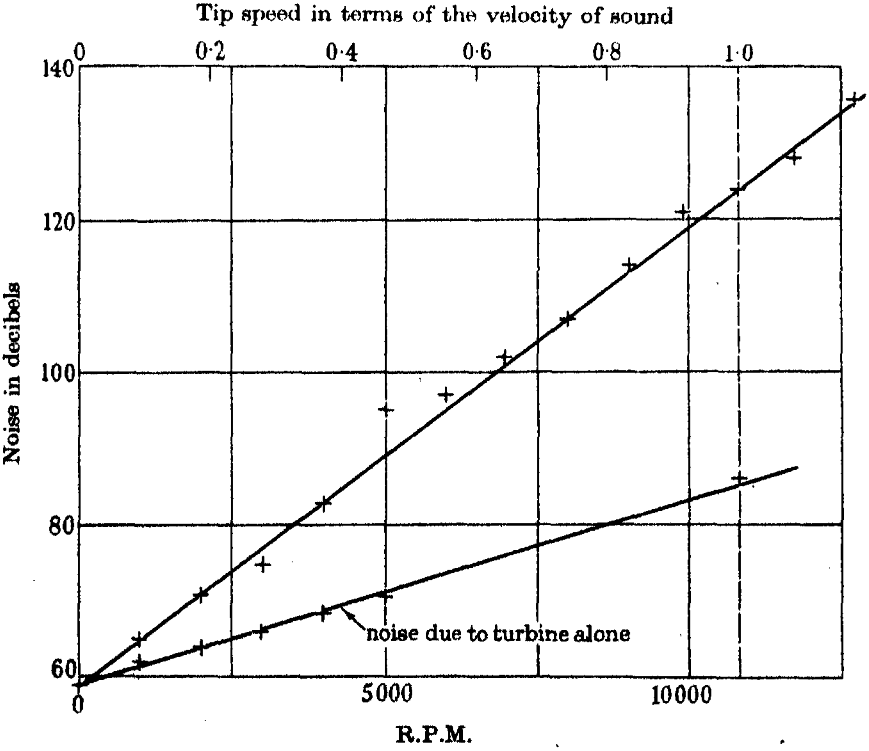


FIG. 5. Variations of loudness with speed, R.A.F. 32 section airscrew with 10·5° incidence at tip. Noise measured in decibels with an audiometer. The noise level of 59 db. at zero speed is due to the air compressor actuating the turbine.

as high as $V = 1{\cdot}2a$. Although there was no discontinuity in loudness as the speed passed through the speed of sound, a very definite change in quality was heard. At lower speeds the note was of the kind associated with a prominent fundamental, accompanied by a few weaker harmonics. On increasing the speed, a change was heard at the precise speed at which shock waves first appeared on the photographs. The note became harsh, and of the type associated with a fundamental accompanied by a large number of prominent harmonics. The existence of shock waves on the photographs suggests a rapid rise in pressure on passing through the wave front, the Fourier analysis of which would be rich in harmonics.

Bearing in mind the continuity of the loudness-R.P.M. curve, it seems that at low speeds the excess pressure in the sound wave is spread out over a band whose width is of the order of half the wave-length, but that at a certain critical speed the wave front suddenly condenses into a single line, retaining the same pressure amplitude as before. Additional evidence was obtained from photographs at this critical speed (see fig. 6). A series of faint wavelets spreads out over a fairly wide area at the critical speed, and condenses into a single wave front at a slightly higher speed.

The formation of a shock wave may be due to small disturbances from the rear of the blade which travel forward with the velocity of sound until they enter a region of sonic velocity, where their motion is arrested by the air moving towards them with equal and opposite velocity. Here they remain fixed relative to the blade until further disturbances arrive to increase the pressure amplitude, whereupon the wave moves forward to a region of even higher velocity, due to its increased rate of propagation. At the critical speed, however, the wave will move forward into a region of lower speed, get ahead of the blade, and become detached from it, in the form of a wavelet. This is precisely what may be seen to be happening in fig. 6, which was taken at the critical speed.

This critical speed at which the sound wave condenses into a single pressure discontinuity is probably the shock stall or compressibility stall, i.e. the speed at which the aerodynamical coefficients show abrupt changes from their low speed values, due to the local velocity of sound being first exceeded at a point on the blade or aerofoil.

The photographs may also be used to throw some light on the generation of airscrew noise by the periodic shedding of eddies from the blade. Measurement of the eddies shows that they are about 1 mm. apart, and if the blade producing them moves with the velocity of sound, then the frequency of shedding eddies will be 340,000 per sec. A sound of this frequency is completely inaudible, and it is improbable that much of the noise actually heard was due to the periodic shedding of eddies from the blades. It may be remarked that if an airscrew blade emits noise continuously, then by virtue of the geometrical construction outlined in § 4, this noise will be concentrated into compressions and rarefactions having the same frequency as the number of airscrew blades passing a given point per second. This holds true even at low subsonic tip speeds.

Equation (11) gives a method of finding the pressure amplitude of shock waves from photographs. It is particularly useful in the case of small amplitude shock waves, or of waves propagated through moving air. Although the method is apparently not very accurate it may be used in cases such as the present which would be otherwise insoluble. The wave shape is fairly satisfactorily explained by theory as is the measured wave velocity.

The well-known linear law connecting noise in decibels with airscrew tip speed has been found to hold up to speeds higher than those used in airscrew design.

In conclusion the author would like to thank the Director of Scientific Research, Air Ministry, for permission to use the high-speed airscrew, members of the staff of the Royal Aircraft Establishment for their valuable practical assistance, Dr Maccoll and the Ordnance Committee, Woolwich Arsenal, for the loan of a bullet photograph, and Professor G. I. Taylor for his helpful criticism of the work.

Finally, his thanks are due to Professor L. Bairstow, under whose direction the research was carried out at the Imperial College of Science and Technology.

SUMMARY

An experimental technique has been developed for photographing the shock waves generated by an airscrew at all speeds above a certain critical speed. A modified form of spark photography was employed, and photographs were obtained for the first time of the airscrew sound wave system at tip speeds ranging from 0·83 to 1·2 of the velocity of sound.

The photographs also reveal a tailing vortex from the blade tip, and an eddying wake behind the blade. These vortices and eddies are more distinct at high speeds than at low, but there is no definite lower limit to the speed at which they can be photographed.

A method is given for calculating the pressure amplitude of these shock waves from measurements of the width of the image. The wave velocity

has been calculated from the photographs, and is found to be slightly greater than that of small amplitude waves.

A method of calculating the shape of the waves is examined, but no theoretical indication of a wave front at tip speeds less than the speed of sound is given, whereas a very definite wave front exists down to speeds as low as 0·85 of sound velocity. The problem of noise reduction in aircraft is discussed.

REFERENCES

Capon 1933 *Engineering*, 27 October.
Douglas and Hartshorn 1931 *Rep. Mem. Aero. Res. Comm., Lond.*, No. 1438.
Glauert 1927 *Proc. Roy. Soc.* A, **118**, 113.
Lamb 1921 *Note on Rep. Mem.* No. 684, *Aero. Res. Comm., Lond.*

KEY TO ANNOTATION OF PLATES 6–11

B = bow wave; R = reflected wave; 1 = first section of wave; 3 = third section of wave; V = vortex from tip; V_2 = tip vortex from previous blade; V_3 = tip vortex from present blade during previous revolution; E = eddying wake; E_2 = eddying wake from previous blade.

The dissociation of the ammonium ion and the basic strength of ammonia in water

By D. H. Everett and W. F. K. Wynne-Jones

(*Communicated by Sir Harold Hartley, F.R.S.—Received* 24 *June* 1938)

In recent years the strengths of many acids have been determined by exact methods and the work of Harned and Embree (1934) on the dissociation constants of acids at different temperatures has given us precise knowledge of the thermodynamic functions associated with the ionization of these substances. No similar measurements had been made with uncharged bases until Pedersen (1937*a*) determined the dissociation constants of aniline and *o*-chloraniline, Pedersen (1937*b*); the accepted values for the common bases are derived from old measurements which can no longer be regarded as sufficiently precise.

The difficulties in the determination of the strength of a base such as ammonia are greater than for an acid such as acetic acid, partly because the effects of carbon dioxide and of the solubility of glass are more serious for alkaline solutions, and partly because the solubility of silver chloride in ammonia prevents the use of the silver-silver chloride electrode. To avoid these difficulties, we determined the hydrogen-ion concentrations of buffer solutions of ammonia and ammonium chloride from measurements of the e.m.f.'s of cells of the type

$$\begin{array}{c|c|c|c|c} \mathrm{H_2} & c_1\,\mathrm{NH_3} & 3\cdot5\mathrm{N} & c\,\mathrm{HCl} & \mathrm{H_2}, \\ & c_2\,\mathrm{NH_4Cl} & \mathrm{KCl} & C-c\,\mathrm{KCl} & \\ & C-c_2\,\mathrm{KCl} & & & \end{array}$$

where the total ion concentration is the same in both half-cells and is equal to $2C$. In a cell of this type the activity coefficient of the hydrogen ion is virtually the same in both compartments and the small liquid junction potential can be completely eliminated by Pedersen's device of making measurements with various ratios of KCl and the other constituents of the solution (i.e. HCl and the buffer mixture) and extrapolating the apparent hydrogen-ion concentrations to zero concentration of the buffer. This extrapolation enables us to obtain the true value of the dissociation constant at each ionic strength and an extrapolation of these values to zero ionic strength gives the thermodynamic dissociation constant. We made measurements at different temperatures so as to obtain values for the other thermodynamic functions.

EXPERIMENTAL

The electrical equipment consisted of a Tinsley "Concentric" potentiometer the scale of which was calibrated by the makers, a high sensitivity galvanometer of 463 ohms resistance, and a Weston standard cell. The apparatus was placed on an equipotential shield which was connected to the thermostats.

The thermostats were controlled nominally at 5°, 15°, 25°, 35° and 45° by relays of the type described by Temple (1936); the relays were actuated by benzene-mercury thermoregulators and the temperatures were kept constant to about 0·01°. Two mercury-in-glass thermometers which could be read to 0·005° and which had recently been calibrated at every fifth degree at the N.P.L. were used for determining the temperatures.

The 15° thermostat was cooled by a current of water passing through a metal coil, while the 5° thermostat was cooled by circulation of brine from a refrigerating unit assembled by Mr P. Temple, for the loan of which we are indebted to Captain J. Golding of the National Institute for Research in Dairying.

The cells were similar to those used for determining the ionic product of heavy water (Wynne-Jones 1936), but they were made entirely of Monax glass and all the ground joints were interchangeable; each half-cell required about 15 c.c. of solution. To saturate the hydrogen with the vapour of the solution two bubblers were placed in front of each half-cell: these bubblers, which are similar to those described by Bichowsky and Storch (1915), were found to be quite efficient in saturating the hydrogen while they produced only a small pressure head and required only 10 c.c. of solution. The solutions were saturated by bubbling at each temperature.

The hydrogen electrodes were made of platinum foil covered with a black deposit according to the instructions of Popoff, Kunz and Snow (1928): if the electrodes differed in potential by more than 0·05 mV they were replatinized.

Hydrogen was obtained from a cylinder and was passed through a tube containing active copper at 450°. Because of the symmetry of the cell it was unnecessary to reduce the results to a pressure of 1 atm. of hydrogen; even at 45° the partial pressure of the ammonia is not sufficient to cause an error of more than 0·05 mV.

The ammonia buffers were prepared by adding definite volumes of standardized carbonate-free KOH solution to a solution of NH_4Cl, sufficient KCl solution was then added to attain the ionic strength required. The acid solutions were prepared from constant-boiling hydrochloric acid (Foulk and Hollingsworth 1923) which was also used for the standardization of the KOH. Kahlbaum's purest ammonium chloride was used except for a few measurements made with a sample of synthetic ammonium chloride: the salt was dried by heating to 130° for 2–3 hr. and then kept in a vacuum desiccator. All the solutions were prepared with conductivity water from a still of the type described by Stuart and Wormwell (1930). All vessels were made of Monax or Jena glass and solutions were used as soon as possible after preparation.

At ammonia concentrations lower than 0·015 N the values of $K_{NH_4^+}$ progressively deviated from the linear relation between log K and the buffer concentration found for the higher concentrations. The direction of the effect could be explained by assuming the removal of ammonia from the solution, but this removal was not the result of the passage of hydrogen through the solution nor of reaction with the glass of the cell and the temperature coefficient of e.m.f. for these cells was unaffected. We believe the explanation to be absorption of carbon dioxide during the preliminary

handling of the solutions. The effect was negligible at concentrations greater than 0·015 N, and in our final measurements we did not employ more dilute solutions.

RESULTS

The actual experimental results are given in Table I of which the first column contains the ionic strength, the second the concentration of ammonia, the third the concentration of NH_4Cl, the fourth that of HCl, the fifth the temperature, the sixth the e.m.f. in volts, and the seventh the temperature coefficient of the e.m.f. The concentration of KCl added to the buffer and to the HCl solution is not given as it can be calculated from the other data. It will be observed that the measurements were not made at round temperatures, but the e.m.f. was found to be very exactly a linear function of the temperature and thus it is possible to interpolate values of the e.m.f. for any desired temperature using the temperature coefficients given in the table.

All concentrations are expressed in g.-mol./1000 c.c. of solution at 15°.

TABLE I

I	c_{NH_3}	c_{NH_4Cl}	c_{HCl}	$t°$	E	dE/dT
0·200	0·0400	0·0400	0·0200	16·70	0·45040	
				25·15	0·44775	
				34·85	0·44490	−0·000304
	0·0394	0·0406	0·0200	5·40	0·45280	
				16·30	0·44980	
				24·75	0·44735	
				33·20	0·44465	
				41·60	0·44210	−0·000304
	0·0295	0·0305	0·0150	5·40	0·44615	
				16·40	0·44260	
				24·70	0·43990	
				33·65	0·43700	
				42·00	0·43400	−0·000322
	0·0236	0·0244	0·0120	5·45	0·44120	
				16·50	0·43700	
				24·75	0·43415	
				33·50	0·43110	
				42·40	0·42780	−0·000347
0·150	0·0400	0·0400	0·0150	5·50	0·44605	
				17·00	0·44235	
				24·80	0·43962	
				34·80	0·43635	
				43·30	0·43340	−0·000330
	0·0320	0·0320	0·0120	5·35	0·44060	
				17·00	0·43675	
				25·25	0·43400	
				34·35	0·43090	
				43·30	0·42730	−0·000348

Table I (*continued*)

I	c_{NH_3}	c_{NH_4Cl}	c_{HCl}	$t°$	E	dE/dT
	0·0250	0·0250	0·00938	5·45	0·43475	
				17·00	0·43065	
				24·75	0·42795	
				34·20	0·42485	
				44·40	0·42080	−0·000354
	0·0200	0·0200	0·0075	17·00	0·42500	
				25·00	0·42185	
				34·80	0·41810	
				44·85	0·41405	−0·000389
0·100	0·0400	0·0400	0·0100	5·40	0·43550	
				16·90	0·43170	
				24·75	0·42855	
				34·75	0·42480	
				44·10	0·42130	−0·000361
	0·0320	0·0320	0·0080	5·40	0·42965	
				16·90	0·42555	
				25·10	0·42270	
				34·80	0·41910	
				44·10	0·41565	−0·000365
	0·0250	0·0250	0·00625	5·50	0·42450	
				17·00	0·41985	
				24·75	0·41685	
				34·80	0·41275	
				44·70	0·40855	−0·000398
	0·0200	0·0200	0·0050	5·55	0·41865	
				17·00	0·41360	
				24·75	0·41040	
				42·30	0·40260	−0·000425
	0·0150	0·0150	0·00375	5·55	0·41200	
				16·80	0·40695	
				24·75	0·40335	
				32·80	0·39985	
				42·30	0·39510	−0·000448
0·070	0·0300	0·0400	0·0070	5·40	0·41975	
				17·00	0·41470	
				24·80	0·41145	
				34·80	0·40725	
				44·90	0·40293	−0·000425
	0·0240	0·0320	0·0056	5·45	0·41455	
				17·00	0·40945	
				24·75	0·40605	
				34·80	0·40170	
				44·90	0·39730	−0·000438
	0·0250	0·0250	0·00438	5·40	0·41520	
				17·00	0·41010	
				24·75	0·40640	
				34·95	0·40160	−0·000425
	0·0200	0·0200	0·00350	5·40	0·41010	
				17·00	0·40500	
				24·80	0·40150	
				34·40	0·39740	
				42·05	0·39385	−0·000438

For each cell the e.m.f.'s given in Table I were plotted against the temperature and the values read off these curves at round temperatures were used to calculate the apparent values of $\log K_{NH_4^+}$ from the equation

$$\log K_{NH_4^+} = \frac{FE}{2{\cdot}303RT} + \log C_{HCl} + \log \frac{c_{NH_3}}{c_{NH_4^+}}.$$

These values are given in Table II which also contains the extrapolated values for zero concentration of the buffer and of HCl. The factor $2{\cdot}303RT/F$ was taken to be 0·05913 at 25° and correspondingly at the other temperatures. The extrapolation to zero concentration of the buffer, which is shown in fig. 1 for the results at 25°, does not seriously affect the values of the dissociation constant, but its importance lies in the fact that it enables the complete elimination of the liquid junction potential.

Table II

I	$c_{NH_4^+}$	$-\log K_{5^\circ}$	$-\log K_{15^\circ}$	$-\log K_{25^\circ}$	$-\log K_{35^\circ}$	$-\log K_{45^\circ}$
0·200	0·0400	9·9288	9·5885	9·2735	8·9777	8·7015
	0·0406	9·9292	9·5895	9·2755	8·9795	8·7036
	0·0305	9·9290	9·5909	9·2769	8·9837	8·7079
	0·0244	9·9302	9·5917	9·2772	8·9873	8·7053
	0	9·933	9·594	9·279	8·985	8·709
0·150	0·0400	9·9131	9·5738	9·2589	8·9636	8·6873
	0·0320	9·9101	9·5746	9·2619	8·9688	8·6954
	0·0250	9·9122	9·5765	9·2641	8·9711	8·6966
	0·0200	9·9141	9·5736	9·2592	8·9649	9·6885
	0	9·915	9·578	9·265	8·973	8·696
0·100	0·0400	9·8980	9·5600	9·2455	8·9517	8·6759
	0·0320	9·8890	9·5536	9·2564	8·9534	8·6812
	0·0250	9·8934	9·5552	9·2425	8·9483	8·6734
	0·0200	9·8937	9·5549	0·2400	8·9448	8·6695
	0·0150	0·8999	9·5616	9·2472	8·9530	8·6765
	0	9·898	9·560	9·248	8·956	8·678
0·070	0·0400	9·8925	9·5537	9·2382	8·9423	8·6658
	0·0320	9·8956	9·5569	9·2426	8·9481	8·6724
	0·0250	9·8900	9·5488	9·2315	8·9346	8·6580
	0·0200	9·8944	9·5576	9·2451	8·9534	8·6796
	0	9·893	9·554	9·238	8·945	8·670

In Table III we have combined these results for the ammonium ion with the data of Harned and his collaborators (1933*a*, *b*; 1935, 1937*a*, *b*) for the ionic product of water, thus obtaining values of the equilibrium constant for the reaction between ammonia and water giving ammonium and hydroxyl ions. Fig. 2 shows the extrapolation of $\log K_{NH_4^+}$ to zero ionic strength.

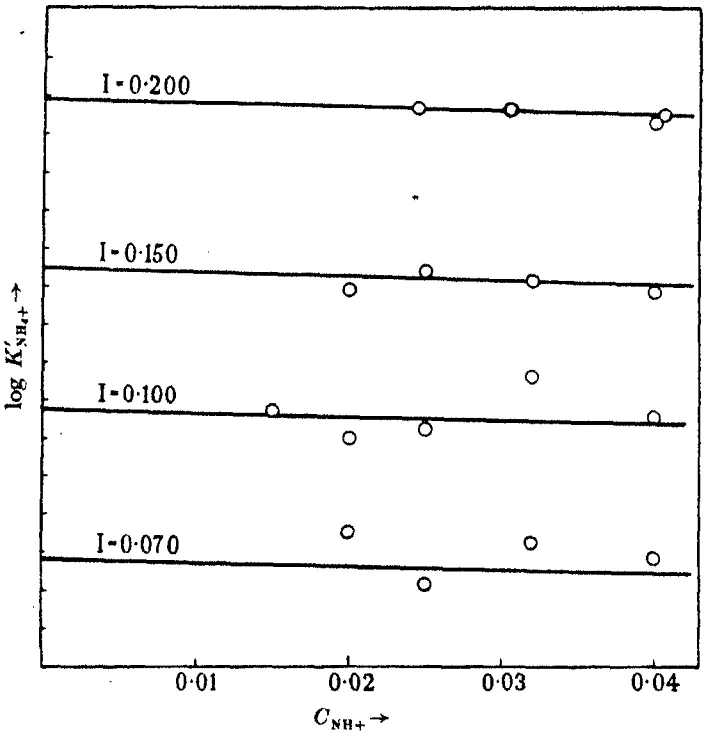


FIG. 1

TABLE III

	I →	0·20	0·15	0·10	0·07	0
5°	$-\log K_{NH_4^+}$	9·933	9·915	9·898	9·893	9·867
	$-\log K_{H_2O}$	14·496	14·512	14·533	14·557	14·734
	$-\log K_{NH_3}$	4·563	4·597	4·635	4·664	4·867
15°	$-\log K_{NH_4^+}$	9·594	9·578	9·560	9·554	9·529
	$-\log K_{H_2O}$	14·107	14·122	14·144	14·166	14·346
	$-\log K_{NH_3}$	4·513	4·544	4·584	4·612	4·817
25°	$-\log K_{NH_4^+}$	9·279	9·265	9·248	9·238	9·215
	$-\log K_{H_2O}$	13·754	13·768	13·791	13·810	13·997
	$-\log K_{NH_3}$	4·475	4·503	4·543	4·572	4·782
35°	$-\log K_{NH_4^+}$	8·985	8·973	8·956	8·945	8·923
	$-\log K_{H_2O}$	13·436	13·450	13·474	13·490	13·680
	$-\log K_{NH_3}$	4·451	4·477	4·518	4·545	4·757
45°	$-\log K_{NH_4^+}$	8·709	8·696	8·678	8·670	8·645
	$-\log K_{H_2O}$	13·148	13·161	13·185	13·204	13·396
	$-\log K_{NH_3}$	4·439	4·465	4·507	4·534	4·751

It is interesting to compare the temperature coefficients of these different equilibrium constants and, as shown in fig. 3, the values of log K for the ammonium ion vary linearly with the reciprocal of the absolute temperature, whereas the values for the ionic product of water and for ammonia do not

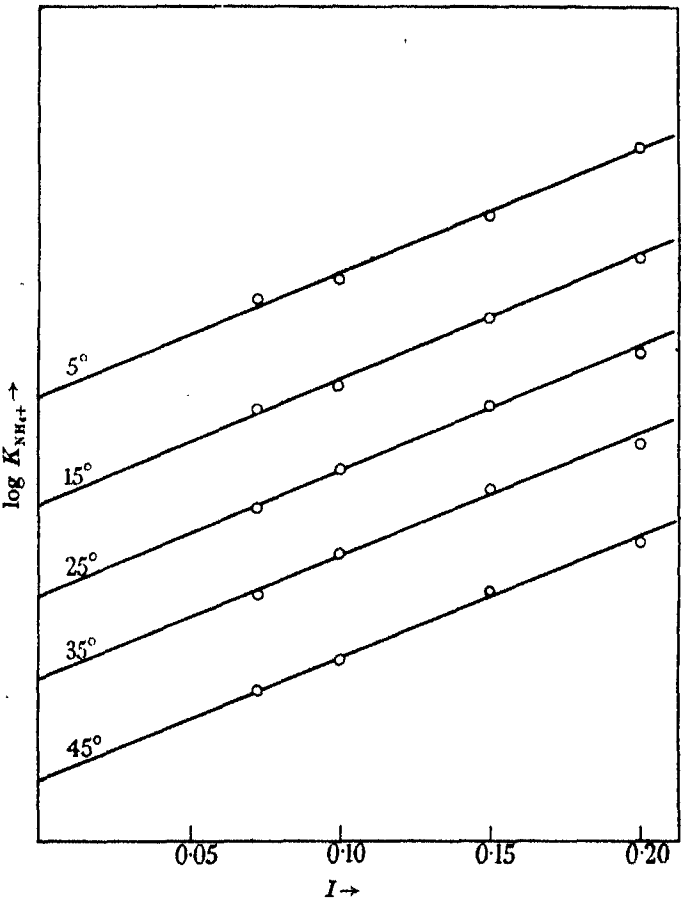


Fig. 2

show this linear relation. The most satisfactory proof of the linearity of the relation between log K and $1/T$ for the ammonium ion is afforded by the plots of e.m.f. against temperature which show that the deviations from a straight line do not exceed two-tenths of a millivolt, i.e. less than 1 % in the dissociation constant. If the e.m.f. changes linearly with the temperature, it follows from the relation between the e.m.f. and the dissociation constant that the latter is related to the temperature by an equation of the form

$$\log K = A/T + B.$$

The temperature coefficient of the e.m.f. of a cell can then be expressed as

$$dE/dT = 2{\cdot}303R/F(B + \log c_{H^+} + \log c_{NH_3} - \log c_{NH_4^+}),$$

a relation which affords a sensitive test of the accuracy of the results, since on plotting dE/dT against $(\log c_{H^+} + \log c_{HN_3} - \log c_{NH_4^+})$ we should obtain not merely a straight line, but a line of slope $2{\cdot}303R/F$. This is strictly true only for constant ionic strength, but the uncertainty in dE/dT is greater than the errors introduced by varying the ionic strength. Fig. 4 represents the temperature coefficients given in Table I.

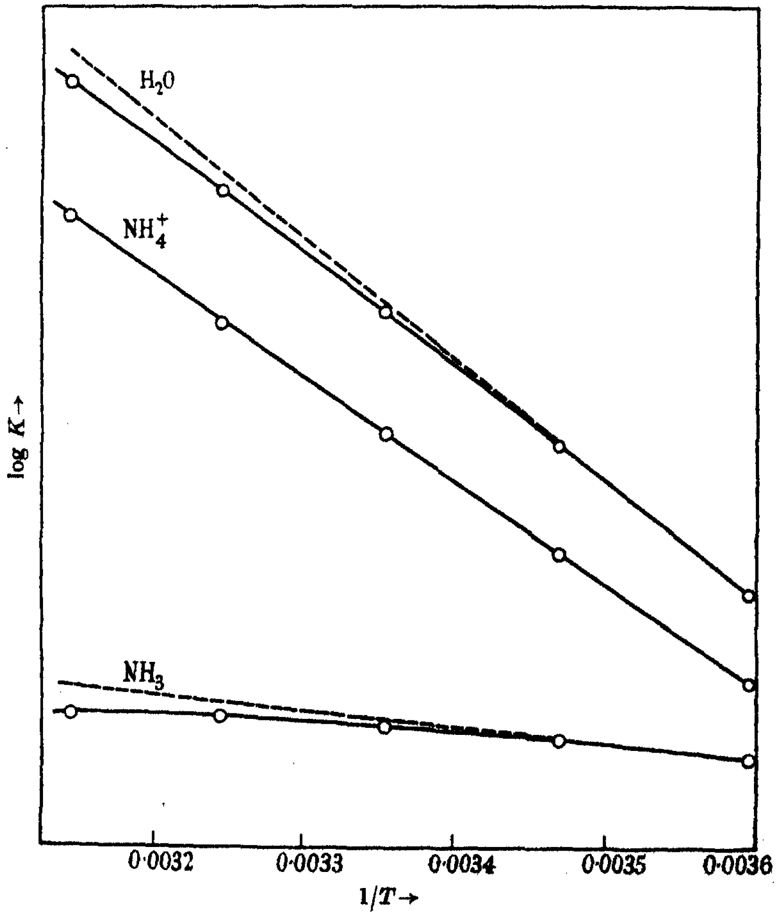


Fig. 3

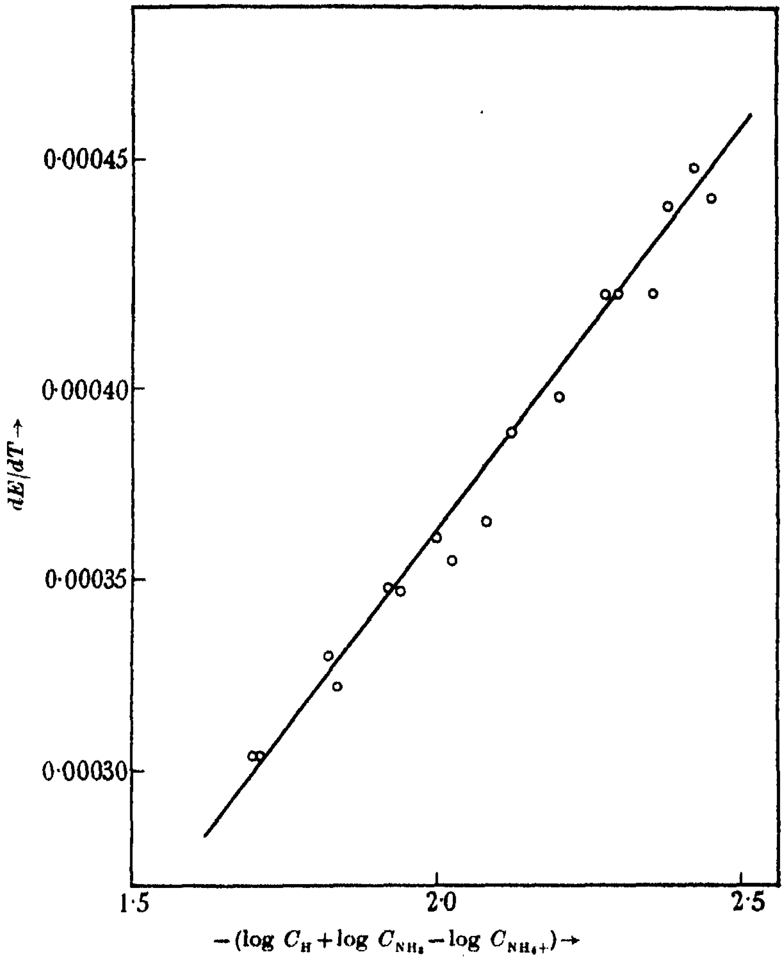


FIG. 4

COMPARISON WITH OTHER MEASUREMENTS

Determinations of the basic strength of ammonia have been made mainly by the electrical conductivity method and there have been numerous investigations dating back to the work of Kohlrausch in 1876. The early measurements gave values greater than 2×10^{-5} for the constant at room temperature for the equilibrium $NH_3 + H_2O = NH_4^+ + OH^-$, but there is no doubt that these results were high because of acid impurities in the water used for making up the solutions. The best of the more recent determinations seem to agree on a value of about $1{\cdot}80 \times 10^{-5}$ at 25°; however, in none of these investigations have the corrections for interionic forces been applied, and for acetic acid, which has a dissociation constant comparable with the basic dissociation constant of ammonia, we know that the application of these corrections has reduced the value of the dissociation constant from $1\cdot85\times10^{-5}$ to $1\cdot75\times10^{-5}$. The only determination which can be regarded as satisfying modern requirements is that of Owen (1934) who, using a hydrogen and a silver-silver iodide electrode in buffer solutions of ammonia and ammonium iodide, obtained a value of $1\cdot75\times10^{-5}$ at 25°: however, this work was reported only in a note and was admittedly a preliminary attempt to see if the silver-silver iodide electrode could be used for determining the strengths of bases. As most of the errors particularly in the conductivity method would increase the value of the dissociation constant and as, in spite of careful analysis, we have been unable to find any serious source of error in our work, we regard $1\cdot65\times10^{-5}$ as the most probable value at 25° for the constant for the equilibrium between ammonia and the solvent water yielding ammonium and hydroxyl ions. We are made more confident of the correctness of this value by its agreement with a series of measurements made by one of us at Princeton University in 1934 with entirely different materials and a procedure which differed from that adopted in this work in sufficient important details to make it improbable that the errors would be the same in the two investigations. The earlier work was carried only at 25° and yielded exactly the same value as that recorded here but with about twice as great a probable error—3 units in the last significant place.

Although we believe that previous determinations of the dissociation constant have given values which are too high, it is probable that in any given work the errors will be roughly the same at each temperature and that, in consequence, the temperature coefficient and heat of ionization derived from these earlier measurements will be essentially correct. In order to test this we have collected in Table IV the results of various investigators on the dissociation constant of ammonia at different temperatures and in fig. 5 we have plotted the values of $\log K + 20\log T$ against $1/T$ after reducing all the constants by a factor which makes them agree at 25° with the value obtained in this work: the term $20\log T$ has been introduced simply because, as shown by Pitzer (1937) it gives a nearly linear representation of the data. The agreement between the various series of measurements is very satisfactory and shows that the temperature coefficient of the equilibrium constant is not seriously affected by the errors which tend to make the absolute values too high.

This conclusion is confirmed by comparison of the heat of dissociation of the ammonium ion calculated from our results with the directly determined

TABLE IV

t° C	Lunden*	Noyes*	Moore and Winmill*	This work
0	$1{\cdot}4 \times 10^{-5}$	1·39	—	—
5	—	—	—	1·358
10	1·63	—	—	—
15	1·70	—	—	1·524
18	1·75	1·72	1·72	—
25	1·87	1·80	1·80	1·652
32·35	—	—	1·89	—
35	—	—	—	1·750
40	1·98	—	—	—
45	—	—	—	1·774
50	1·96	1·81	—	—
60	1·9	—	—	—
75	—	1·64	—	—
100	—	1·35	—	—

* Lunden (1907); Noyes (1907); Moore and Winmill (1912).

values of Pitzer for the heats of ionization of ammonia and water. Our value of $\Delta H_{\mathrm{NH_4^+}}$ is 12,400 cal. and the value deduced from Pitzer's data is 12,493 cal. The agreement is within the experimental errors of the determinations.

The dissociation constant of an electrolyte is related to the heat of ionization by the well-known equation

$$\frac{d \ln K}{dT} = \frac{\Delta H}{RT^2},$$

and we also have Kirchhoff's relation between the temperature coefficient of the heat of reaction and the difference in the heat capacities of the reactants and products of reaction

$$\frac{d(\Delta H)}{dT} = \Delta C_p.$$

If ΔC_p is regarded as constant, combination of these two equations gives us the equation

$$\log K = \frac{A}{T} + \frac{\Delta C_p \log T}{R} + B.$$

which has been shown by Pitzer to represent Harned's data on the dissociation of uncharged acids remarkably well if ΔC_p is assigned a value of -40 cal. per degree. On the other hand, the results of the present work show that for the dissociation of an acid carrying a single positive charge the value of ΔC_p is not appreciably different from 0. This striking difference in the behaviour of these two types of acid may be regarded as evidence that the value of ΔC_p is determined primarily by electrostatic effects, the heat capacity of an ionic species being markedly less than that of an ordinary uncharged molecule. Since it has been shown that in similar solvents certain types of ionic equilibria depend upon the dielectric constant in the way expected from the Born equation (Wynne-Jones 1933), i.e.

$$\Delta \ln K = \frac{\epsilon^2}{kTr} \cdot \Delta\left(\frac{1}{D}\right),$$

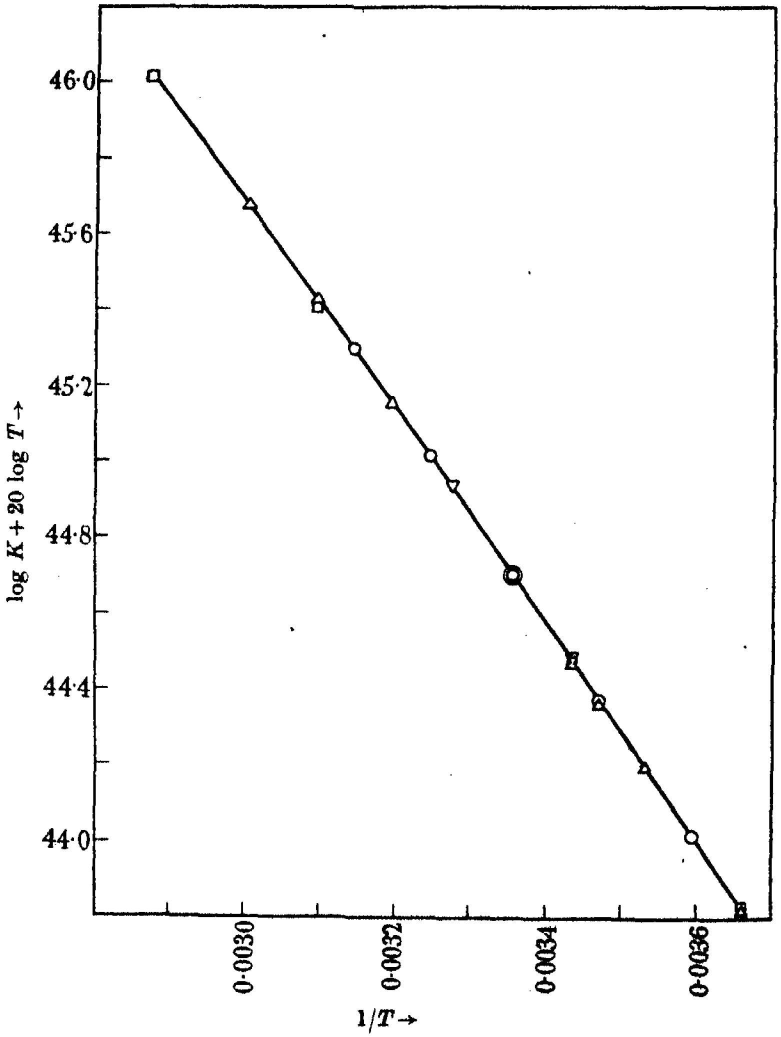


Fig. 5. ⊙ This work, △ Lunden, ▽ Moore and Winmill, ⊡ Noyes, ◉ 25° point of this work. The other series have all been reduced to agree with this value.

we might expect it to be possible also to treat the temperature dependence in the same manner. We find, however, that if we assume $r = 1$ A the value of ΔC_p calculated for an ionization reaction in water cannot be numerically greater than -7 cal. per degree. This difference can be reconciled with the view that the effect is electrostatic in nature by postulating such powerful orientation of solvent molecules by the fields of the ions that the contribution of the rotational degrees of freedom of these molecules to the heat capacity has been lost. This hypothesis has important consequences for the applicability of the Born equation to the change of equilibrium constant with solvent and it will be interesting to determine temperature coefficients in different solvents. It would also appear that the net charge on a molecule may be of less importance than the possibility of orienting a large number of solvent molecules, a conclusion which is supported by the recent work of Smith, Taylor and Smith (1937) on the ionization of amino acids. Examination of the results of these workers for 8 α-amino acids show that ΔC_p for the first dissociation stage is 36 cal. per degree, while for the second stage it is about 14 cal. per degree. Now the two stages are represented by the equations

$$CH_2NH_3^+COOH = CH_2NH_2COOH + H^+,$$
$$CH_2NH_2COOH = CH_2NH_2COO^- + H^+,$$

and, at first sight, we might expect the respective values of ΔC_p to be approximately 0 and -40. However, it has been shown by Adams (1916) and Bjerrum (1922) that the molecules of amino acids are really ampho-ions, $CH_2NH_3^+COO^-$, and the first stage of dissociation should be represented as

$$CH_2NH_3^+COOH = CH_2NH_3^+COO^- + H^+,$$

while the second stage is

$$CH_2NH_3^+COO^- = CH_2NH_2COO^- + H^+.$$

Regarded in this way the first stage corresponds more closely to the ionization of an uncharged acid, since the number of charge centres increases by two when the reaction proceeds from left to right. On the other hand, in the second stage the number of charge centres remains constant, and consequently we should expect the dissociation constant to simulate the behaviour of a positively charged acid. Of course, it is not to be expected that the ampho-ion will behave as two quite independent charges, but the experimental values of the heat capacities show that there is rather strong orientation of solvent molecules around both the charges of the ampholyte.

We are indebted to Dr G. Salomon for his help in making some of the preliminary measurements and to the Government Grant Committee of the Royal Society for the loan of the potentiometer.

SUMMARY

1. The dissociation of the ammonium ion in water has been measured at 10° intervals from 5° to 45°. The results can be represented by the equation

$$-\log K_{\mathrm{NH}_4^+} = \frac{2706}{T} + 0\cdot139,$$

2. Values of the constant for the equilibrium $\mathrm{NH}_3 + \mathrm{H}_2\mathrm{O} = \mathrm{NH}_4^+ + \mathrm{OH}^-$ have been derived from these results by combining them with the data for the ionic product of water. At 25° K_{NH_3} is $1\cdot65 \times 10^{-5}$.

3. The heat of dissociation of the ammonium ion as calculated from the temperature coefficient of the dissociation constant is 12,400 cal.

4. The temperature dependence of dissociation constants is related to the electrical type of the chemical reaction and, from the fact that simple electrostatic considerations do not account for the results, it is concluded that strong orientation of the solvent molecules occurs.

REFERENCES

Adams 1916 *J. Amer. Chem. Soc.* **38**, 1503.
Bichowsky and Storch 1915 *J. Amer. Chem. Soc.* **37**, 2696.
Bjerrum 1922 *Z. phys. Chem.* **104**, 147.
Foulk and Hollingsworth 1923 *J. Amer. Chem. Soc.* **45**, 1220.
Harned and Copson 1933*b* *J. Amer. Chem. Soc.* **55**, 4496.
Harned and Donelson 1937*a* *J. Amer. Chem. Soc.* **59**, 1280.
Harned and Embree 1934 *J. Amer. Chem. Soc.* **56**, 1050.
Harned and Geary 1937*b* *J. Amer. Chem. Soc.* **59**, 2032.
Harned and Hamer 1933*a* *J. Amer. Chem. Soc.* **55**, 2194.
Harned and Mannweiler 1935 *J. Amer. Chem. Soc.* **57**, 1873.
Lunden 1907 *J. Chim. phys.* **5**, 574.
Moore and Winmill 1912 *J. Chem. Soc.* **101**, 1635.
Noyes 1907 *Publ. Carneg. Instn*, no. 63.
Owen 1934 *J. Amer. Chem. Soc.* **56**, 2785.
Pedersen 1937*a* *K. danske vidensk. Selsk. Skr.* **14**, 9.
— 1937*b* *K. danske vidensk. Selsk. Skr.* **15**, 3.
Pitzer 1937 *J. Amer. Chem. Soc.* **59**, 2365.
Popoff, Kunz and Snow 1928 *J. Phys. Chem.* **32**, 1056.
Smith, Taylor and Smith 1937 *J. Biol. Chem.* **122**, 109.
Stuart and Wormwell 1930 *J. Chem. Soc.* p. 85.
Temple 1936 *J. Sci. Instrum.* **13**, 414.
Wynne-Jones 1933 *Proc. Roy. Soc.* A, **140**, 440.
— 1936 *Trans. Faraday Soc.* **32**, 1397.

The low temperature properties of gaseous helium

By H. S. W. Massey, Ph.D., and R. A. Buckingham, Ph.D.

(See volume 168, pages 378–389, 7 November 1938)

Errata

Equation (7) should read:

$$\text{``}\, V(r) = 10^{-9} e^{-r/0\cdot 217} - 1\cdot 91 \times 10^{-12} r^{-6} \text{ ergs,} \qquad (7)$$

being 1·30 times that given by Slater. This function, which necessarily has a zero in the same position as Slater's function, provides a suitable starting-point from which to determine the exact interaction, as there is evidence that Slater's function gives too small a repulsion at small distances and too small an attraction at intermediate distances."

This modification affects the discussion in the last paragraph of §3. The differences between the results of this paper and those of Massey and Mohr using the Slater interaction are probably largely due to the difference in magnitude of the interactions employed but may be partly due also to the inaccuracy of the perturbation method employed by Massey and Mohr to obtain the smaller positive phases. The discussion of §3 then applies.

The first sentence of the Summary should read:

"The viscosity and second virial coefficients of helium gas are calculated in the temperature range 0–26° K using exact quantum formulation and an atomic interaction 1·30 times that given by Slater."

The kinetics of electrode reactions. I and II

By J. N. Agar and F. P. Bowden

Laboratory of Physical Chemistry, Cambridge

(*Communicated by R. G. W. Norrish, F.R.S.—Received* 28 *June* 1938)

[Plate 9]

PART I. THE EFFECT OF LOCAL CONCENTRATION CHANGES ON OVERPOTENTIAL

In any surface reaction taking place in a solution, it is clear that the concentration of the reactants in the vicinity of the surface must fall. If the concentration in the bulk of the solution remains constant, a steady state may finally be reached, in which the rate of replenishment of the solution in this region, from the bulk, is equal to the rate at which the reactant in question is used up. But, in general, such a state is only attained when the concentration at the surface is less than that in the rest of the solution. If the reaction considered is an electrode reaction, these concentration changes may affect the electrode potential. This question is therefore of importance in the study of overpotential, and of the kinetics of electrode processes generally.

The overpotential at an electrode is defined as the potential difference between this electrode and a similar unpolarized reversible electrode in the same solution. In practice this reference electrode is usually situated outside the region affected by the concentration changes near the electrode at which the reaction is taking place. The measured potential difference between the two electrodes, i.e. the measured overpotential, may therefore include a term due to the concentration difference.

It is possible to distinguish three main causes of irreversibility in electrode reactions:

(i) Concentration changes of the type here considered. The term "concentration polarization" has often been used in the past to designate irreversibility of this kind. In order to avoid the general confusion which arises from the loose use of the word "polarization", however, we have introduced the expression "concentration overpotential".

(ii) If the surface reaction itself has a large energy of activation, it may be slow compared with other stages of the deposition process, and give rise to an overpotential of a different type, which we have called "activation overpotential". The hydrogen and oxygen overpotentials are of this kind.

(iii) If the electrolyte near the electrode has an appreciable resistance, or if a poorly conducting film is present on the surface, the passage of current will give rise to a potential difference which obeys Ohm's law. We may call this third type of irreversibility "resistance overpotential".

In Part I a general theory of concentration overpotential is developed. The treatment is restricted to the systems in which activation overpotential is small, and in which the electrode process itself can be regarded as reversible. This classification includes the deposition of the halogens (Keenan, unpublished work) and many oxidation-reduction processes.

Part II describes an experimental investigation of oxygen deposition from fused sodium hydroxide. It will be shown that the observed overpotential is due to concentration changes, and that the results are consistent with the general theory.

CONCENTRATION CHANGES IN THE VICINITY OF THE ELECTRODE

When a steady state is reached, it is clear that the rate of removal of a constituent of the solution by the reaction at the electrode must be equal to the rate at which it is brought up from the bulk of the solution, where its concentration is supposed to remain constant. Three processes are of importance in this connexion:

(i) *Migration*, i.e. the movement of ions towards the surface under the influence of an electric field.

(ii) *Convection*, i.e. bodily movement of liquid containing the substance of interest in solution. The movement of the liquid may be due either to mechanical stirring or to "convection currents", arising from density differences between one part of the solution and another.

(iii) *Diffusion*.

In the most general case the calculation of the amount of material transported by migration is a matter of some complexity. In practice, however, it is only necessary to consider certain special cases:

(*a*) The reactant is not an ion (e.g. dissolved chlorine). There is no transport by migration.

(*b*) The reactant is ionized, but the solution contains an excess of some foreign electrolyte, which takes no part in the reaction. In this case transport of the reacting substances by migration may be negligibly small.

(*c*) The solution contains only two species of ions, one of which is a reactant. In this case the transport number of an ion is practically independent of the concentration.

It is difficult to give a complete solution of the hydrodynamical problems involved in the convection and diffusion process. It is therefore necessary to proceed by semi-empirical and approximate methods. In this connexion the concept of the "diffusion layer" has been found very useful.

There can be no motion of fluid across the electrode surface itself, and therefore convection cannot take place at any point on this surface. On the other hand, in the bulk of the solution, at sufficiently large distances from the electrode, the concentration is practically uniform. Diffusion must therefore be of very little importance in this region. Adopting the ideas first put forward by Noyes and Whitney (1897), and developed by Nernst (1904) and Brunner (1904), we may suppose that the electrode surface is covered by a "diffusion layer" of thickness δ. Outside this layer the concentration is constant, and is that of the bulk of the solution. We assume that convection is negligible inside the layer, and that the reactant is supplied to the electrode by diffusion (and migration) alone.

It follows that in the steady state the concentration gradient in the diffusion layer on a plane surface is linear. This state of affairs is represented by *ABC* in fig. 1, which shows the variation of concentration with distance from the surface. The *true* concentration near the surface is probably more accurately represented by the curve *AXC*. It is evident that the rate at which a substance diffuses inwards through the layer is equal to

$$\frac{k(c_0 - c_e)}{\delta}, \tag{1}$$

where k is the diffusion coefficient of the substance, c_0 the concentration in the bulk, and c_e that at the surface, expressed in g.-mols. per c.c. This treatment is equally applicable to the removal of a product of the reaction from the electrode region.

Many calculations of the thickness of the diffusion layer in various systems have been made. A general discussion of most of the experimental data available has recently been given by Glasstone and Hickling (1935). A theoretical treatment of the subject from the standpoint of the theory of dimensions has been carried out by one of the present authors, and will be published shortly.

The theory developed shows that:

(*a*) In *stirred* solutions, δ is independent of the difference in concentration of the bulk solution and that at the surface, and is also very nearly independent of the nature of the solute in all ordinary cases.

(*b*) In *unstirred* solutions, δ is not entirely independent of the concentration difference, but only varies slightly with it. It is again practically independent of the nature of the solute.

This is in agreement with experimental work.

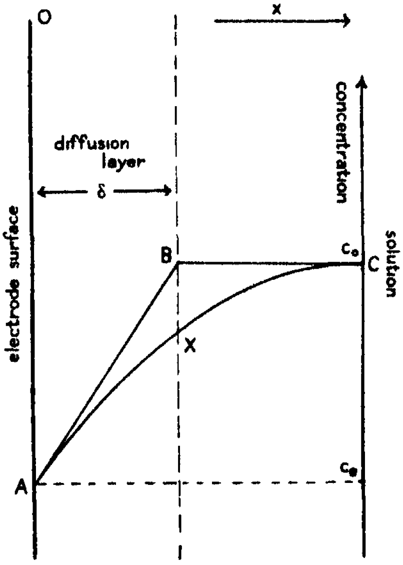


FIG. 1. The diffusion layer.

THE RELATION BETWEEN CURRENT DENSITY AND CONCENTRATION

Consider the deposition of an n-valent metal from a solution containing its ions, and also a large excess of some other electrolyte, e.g. KCl, which takes no part in the electrode reaction. Under these conditions migration can be neglected.

If the current density is I, the number of g.-atoms of the metal deposited, and therefore removed from the solution, per sec. per sq. cm. of the surface, is equal to I/nF, where F represents the faraday.

In the steady state this expression must be equal to the rate of diffusion through the diffusion layer, given by (1). Hence:

$$\frac{I}{nF} = \frac{k(c_0 - c_e)}{\delta}. \qquad (2)$$

It is clear that c_e cannot be less than zero, so that in any given solution the rate of deposition cannot exceed kc_0/δ. The corresponding "limiting current density", I_0, is given by

$$|I_0| = \frac{nFkc_0}{\delta}. \tag{3}$$

If the current density is increased above this value, an alternative electrode process (e.g. liberation of hydrogen) must commence. This is usually associated with a sharp rise in the electrode potential.

We know that in stirred solutions δ is independent of the rate of diffusion through the layer. Consequently, from equations (2) and (3), we have

$$\frac{c_e}{c_0} = 1 - \frac{\delta I}{knFc_0} = 1 - \left|\frac{I}{I_0}\right|. \tag{4}$$

This is also approximately true for unstirred solutions.

In the case here considered the concentration at the interface is less than that elsewhere. In the anodic dissolution of the same metal, on the other hand, it is evident that the concentration at the surface is greater than that in the bulk, so that

$$\frac{c_e}{c_0} = 1 + \left|\frac{I}{I_0}\right|. \tag{5}$$

Finally, we may take migration into account. We suppose that the solution contains only a salt, AB, of the metal, A, being deposited.

If t_A is the transport number and n_A the valency of the ion A, then, *owing to the migration* of these ions, $t_A I/F$ g.-equiv., or $t_A I/n_A F$ g.-mol. pass through unit area of the diffusion layer per second. Migration and diffusion must together transport the ions through the diffusion layer as fast as they are deposited, so that

$$\frac{k(c_0 - c_e)}{\delta} + \frac{t_A I}{n_A F} = \frac{I}{n_A F}. \tag{6}$$

If t_B is the transport number of the anion, B, we have $t_B = 1 - t_A$ and

$$\frac{c_e}{c_0} = 1 - \frac{t_B}{kn_A Fc_0} I = 1 - \left|\frac{I}{I_0}\right|, \tag{7}$$

where

$$|I_0| = \frac{kn_A Fc_0}{\delta t_B}. \tag{8}$$

Comparison with (3) shows that migration effectively increases I_0 by a factor of $1/t_B$.

As mentioned above, the treatment of migration in the more general cases becomes very complicated, since the transport numbers of the ions are no longer independent of the concentration. From the experimental point of view this limitation is not serious.

It is not difficult to extend the equations for the "no migration" case so that they apply to a generalized electrode reaction (anodic or cathodic):

$$a_1 A_1 + a_2 A_2 + \dots + a_i A_i + \dots \nu\epsilon = 0$$

(ϵ represents an electron).

We adopt the following sign convention: a_1, a_2, ..., ν, are taken to be positive for the products of the reaction, and negative for the reactants. The current density, I, is taken to be positive when the current flows from the electrode to the solution. We now have

$$\frac{c_e}{c_0} = 1 - \frac{I}{(I_0)_i}, \tag{9}$$

where

$$(I_0)_i = -\nu F\left(\frac{kc_0}{a\delta}\right)_i. \tag{10}$$

It is of interest to estimate the numerical values of I_0. In the simplest cases

$$a = \nu = n = \pm 1,$$

so that

$$I_0 = \pm\frac{kFc_0}{\delta},$$

where c_0 is expressed in g.-mol./c.c. In unstirred solutions we may take $\delta = 0{\cdot}05$ cm. (Glasstone and Hickling 1935). For most substances, $k \risingdotseq 10^{-5}$ cm.^{2} sec.$^{-1}$, so that $I_0 \risingdotseq \pm 20c_0$ amp./sq. cm. In vigorously stirred solutions I_0 will be at least 10 times greater.

THE VARIATION OF ELECTRODE POTENTIAL WITH CURRENT DENSITY

Provided "activation overpotential" and the potential drop due to the ohmic resistance of the electrolyte can be neglected, the total observed overpotential must be "concentration overpotential", which we represent by the symbol η_c. This potential is clearly equal to that of a concentration cell, in which the concentrations at the two electrodes are c_e and c_0. In cases in which the transport of substances by migration is unimportant, we may neglect the liquid junction potential.

In a simple system we thus have

$$\eta_c = \frac{RT}{nF} \ln \frac{c_\epsilon}{c_0} \tag{11}$$

approximately. We here assume the use of the concentrations c_e and c_0 in place of the corresponding activities will give results of sufficient accuracy. (A system in which this approximation is not adequate is discussed in Part II.)

Taking η_c to be positive when the polarized electrode is positive with respect to the unpolarized electrode, we obtain

$$\eta_c = \frac{RT}{nF} \ln \left(1 - \frac{I}{I_0}\right), \tag{12}$$

after substituting in (11) the value of c_e/c_0 given by (3) or (10).

Corresponding to the general equation (9), we have

$$\begin{aligned}\eta_c &= \frac{RT}{\nu F} \sum_i a_i \ln \left(\frac{c_e}{c_0}\right)_i \\ &= \frac{RT}{\nu F} \sum_i a_i \ln \left\{1 - \frac{I}{I_0}\right\}_i.\end{aligned} \tag{13}$$

In cases in which migration is important, the relevant expression for the e.m.f. of a concentration cell with transport must be used in place of that employed above. This leads to an equation of the same form as (13).

In figs. 2*a* and 2*b* the value of η_c given by equation (12) is plotted against I and $\log_{10} I$, respectively. I_0 is taken to be -10^{-3} amp./sq. cm., since this would be a typical value in approximately 0·1 M solutions.

These curves show the type of current-potential relationship which we should expect when the overpotential is determined by the concentration changes of *only one* of the substances taking part in the reaction. In practice, this condition is often approximately fulfilled. It should be noted that the nature of the relation between overpotential and current density for electrodeposition (I negative) is entirely different from that for electrosolution (I positive). This difference is due to the fact that the potential determining substance is being *used up* at the electrode when I is negative, and is being *produced* when it is positive. In the former case the concentration at the surface cannot *fall* below zero; in the latter it can *increase* practically without limit. For a given numerical value of I, the overpotential is much greater

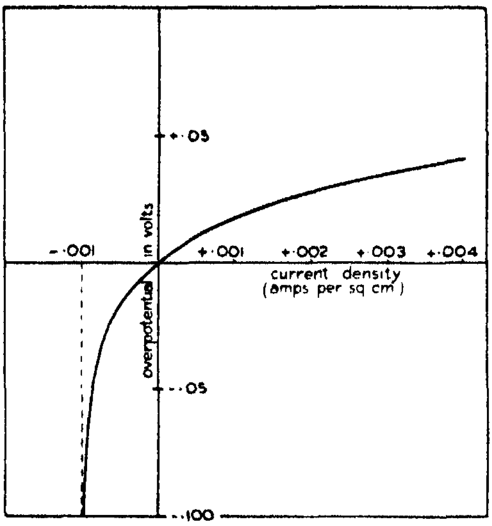


FIG. 2*a*. Relation between concentration overpotential, η, and current density, I.

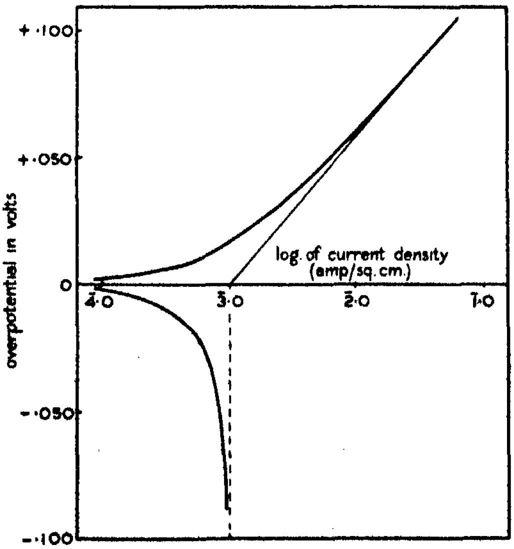


FIG. 2*b*. Relation between concentration overpotential and logarithm of current density.

for negative I than for positive I. On the positive side, however, it is interesting to note that (12) reduces to the linear relation:

$$\eta_c = \frac{RT}{F}\ln\left(\frac{-I}{I_0}\right) \tag{14}$$

between η_c and $\ln I$, for sufficiently large positive values of I. Such a linear relation is commonly associated with activation overpotential.

THE RATE OF GROWTH OF CONCENTRATION OVERPOTENTIAL

The influence of time on the concentration near an electrode was investigated theoretically by Sand (1901), for a system in which convection did not occur. A more general treatment, based on the idea of the diffusion layer, has been given by Roseburgh and Lash Miller (1910). In cases in which migration can be neglected they find

$$\frac{k(c_e - c_0)\nu F}{aI} = 1 - \xi, \tag{15}$$

where $$\xi = \frac{8}{\pi^2}\sum_{m=1,3,5,\ldots}\frac{1}{m^2}e^{-m^2\mu t},$$

the "time constant", $$\mu = \frac{\pi^2 k}{4\delta^2},$$

and m represents odd integers. For sufficiently large values of t, $\xi \to 0$, and (15) reduces to the steady state equation (9). Convenient approximations for ξ are available for either small or large values of μt.

We may now calculate how the concentration overpotential depends on time. We may make the assumption that the quantity of electricity which must be passed in order to produce the concentration changes, and thereby build up the concentration overpotential, is much larger than the corresponding change in the charge of the double layer on the electrode. The latter quantity can be calculated from the known capacity of the double layer (approx. 20 μF/sq. cm. of true surface); direct experiment (Part II) shows that it is usually negligible. In some cases, however, it may be appreciable, and an allowance for it should be made. Neglecting the double layer capacity we can at once calculate the rate of growth of η_c from equation (15).

It is clear that $$\frac{c_e}{c_0} = 1 - (1 - \xi)\frac{I}{I_0}, \tag{16}$$

where I_0 has its previous significance (equation (10)). In the most commonly occurring case, in which the potential is effectively determined by the concentration changes of only *one* substance, we thus have

$$\eta_c = \frac{RT}{\nu F} a \ln\frac{c_e}{c_0} = \frac{RT}{\nu F} a \ln\left\{1-(1-\xi)\frac{I}{I_0}\right\}. \tag{17}$$

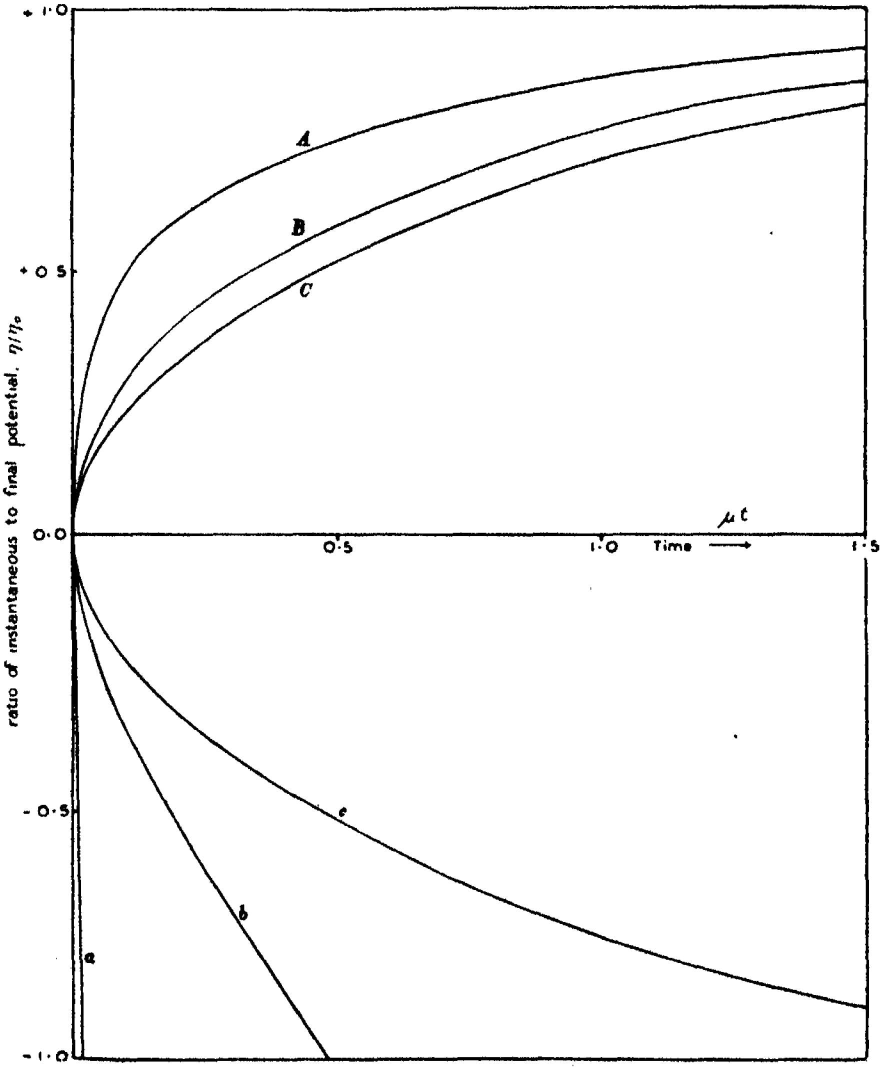


FIG. 3. Rate of growth of concentration overpotential at various current densities.

Curve A $I = -10I_0$
B $I = -I_0$
C $I = -0{\cdot}1I_0$

Curve a $I = +10I_0$
b $I = +I_0$
c $I = +0{\cdot}1I_0$

In the simple case where $n=1$ and $a=1$,

$$\eta_c = -\frac{RT}{F}\ln\left\{1-(1-\xi)\frac{I}{I_0}\right\}.$$

Fig. 3 shows this relation between η_c and μt for various currents which may occur in practice. For convenience, the ratio η/η_0 is plotted in this figure; η_0 is equal to

$$\frac{RT}{F}\ln\left\{1+\left|\frac{I}{I_0}\right|\right\}$$

Taking $k = 10^{-5}$ cm.²/sec., the value of μ is $\dfrac{2\cdot5\times10^{-5}}{\delta^2}$ sec.⁻¹. Table I shows the values of μ corresponding to various values of δ.

From these calculations, we may see how the electrode potential varies with time. The experiments described in Part II show that the relation between concentration overpotential and time is given by these expressions, and that μ is of the right order.

TABLE I

δ cm.	μ sec.$^{-1}$
0·05	0·01
0·01	0·25
0·005	1·0
0·001	25

THE RESISTANCE OF THE SOLUTION NEAR THE SURFACE

The electrode potential measured while the current is flowing will include a potential drop due to the resistance of the solution near the electrode. The magnitude of this is clearly dependent on the experimental set-up and on the conductivity of the solution. The resistance due to oxide or other poorly conducting films on the electrode surface is ignored here.

If the total concentration of electrolytes near the electrode remains constant, this resistive potential drop, or "resistance overpotential", η_r, must be given by an expression of the form

$$\eta_r = \text{const.}\, I.$$

Such an expression is applicable to systems of the "no migration" type, since we assume that in such cases the solution contains a large amount of some electrolyte, in practically uniform concentration throughout the liquid. Under these conditions, it is possible to apply a correction for the resistance overpotential, or, in many instances, to neglect it entirely, in comparison with the concentration overpotential.

But in other systems, e.g. the deposition of a metal from a solution containing one of its salts alone, this is no longer true. Owing to the concentration changes, the resistance of the solution near the electrode is not constant, and may attain very large values if the diffusion layer becomes seriously depleted of ions.

In the present treatment we make the following assumptions:

(*a*) The variation of concentration near the surface is of the linear type illustrated in fig. 1 (line ABC).

(*b*) The whole thickness of the diffusion layer is included in the current path between the electrode and the reference electrode.

For simplicity we restrict ourselves to the case of deposition of the cation A, from a solution of a uni-univalent electrolyte, AB. t_A, t_B are the transport numbers of the A and B ions respectively. In the case of a large plane electrode we then have

$$\eta_r = \text{const.}\, I + \int_0^\delta \frac{I\,dx}{\Lambda c_x}, \tag{18}$$

where Λ is the equivalent conductivity of the electrolyte, and c_x is the concentration at distance x from the electrode. According to assumption (*a*) above

$$c_x = c_e + \frac{x}{\delta}(c_0 - c_e) \quad \text{(see fig. 1).}$$

Introducing the value of I from equation (7) and assuming Λ to be constant, we find that

$$\eta_r = \text{const.}\, I + \frac{Fk}{\Lambda t_B} \ln\left\{1 - \frac{I}{I_0}\right\}, \tag{19}$$

where $$I_0 = -\frac{Fkc_0}{\delta t_B} \quad \text{(cf. (8)).}$$

We may reasonably assume that the diffusion coefficient of an ion is proportional to its mobility. On this basis Nernst (1888) and Haskell (1908) have shown that the diffusion coefficient of an electrolyte, k, is related to the equivalent conductivity Λ, by the equation

$$k = \frac{2RT}{F^2} t_A t_B \Lambda, \tag{20}$$

so that $$\eta_r = \text{const.}\, I + \frac{RT}{F} 2t_A \ln\left(1 - \frac{I}{I_0}\right). \tag{21}$$

In this equation the second term on the right-hand side represents the ohmic potential drop due to the (variable) resistance of the diffusion layer. The first term is the potential drop in that part of the unchanged solution outside the diffusion layer which is included in the potential measuring circuit. *Comparison with* (12) *shows that the ohmic potential drop in the diffusion layer is of the same order, and varies with the current density in the same manner, as the concentration overpotential proper.* This rather striking result is, of course, a consequence of the relation between the diffusion coefficient of an ion and its mobility.

CONCLUSION

In the investigation of the mechanism of irreversible electrode processes, it is desirable to know whether an observed overpotential is of the concentration or activation type. The characteristics of concentration overpotential may be determined from the general theory developed in this paper. Those of activation overpotential are known, though not with complete certainty in all cases, from experimental work carried out on the hydrogen and oxygen overpotentials (e.g. Bowden 1928, 1929; Volmer and Wick 1935; Wirtz 1938).

Some of the differences between the two types of overpotential are set out below.

(1) *The effect of stirring.* The numerical value of the limiting current, I_0, is always increased by stirring, so that the concentration overpotential will fall. Stirring has little or no effect on activation overpotential.

(2) *The nature of the surface.* Activation overpotential is profoundly affected by the nature and physical state of the electrode surface. Concentration overpotential is in general unaffected by these factors. If, however, the surface irregularities are very large, so that they penetrate beyond the diffusion layer, it is possible that they may cause small variations on the concentration overpotential.

(3) *The temperature coefficient.* Differentiation of (12) shows that

$$\left(\frac{\partial \eta_c}{\partial T}\right)_{I=\text{const.}} = \frac{\eta_c}{T} - \frac{RT}{F}\frac{I}{(I-I_0)}\left(\frac{\partial \ln I_0}{\partial T}\right)$$

$$\doteq \frac{\eta_c}{T} - \frac{RT}{F}\left(\frac{\partial \ln I_0}{\partial T}\right)$$

if $$I \gg -I_0.$$

As the temperature is raised, the diffusion coefficient increases, and at the same time the thickness of the diffusion layer may decrease. Hence $\frac{\partial \ln I_0}{\partial T}$ is always positive. A survey of the data indicates that $\frac{\partial \ln I_0}{\partial T}$ is approximately 0·04 per degree in a very large number of cases.

As a typical example we take $\eta_c = 0{\cdot}150$ V and $T = 300°$ K, so that

$$\left(\frac{\partial \eta_c}{\partial T}\right)_I = 0{\cdot}0005 - 0{\cdot}001 = -0{\cdot}5 \times 10^{-3}\,\text{V/degree.}$$

Under similar conditions the temperature coefficient of activation overpotential is considerably higher—about $-2{\cdot}5 \times 10^{-3}$ V/degree.

(4) *The effect of the concentration of the solution.* Apart from effects of secondary importance, I_0 is directly proportional to the concentration of the important substance in the solution. Although a similar relation between concentration and *activation* overpotential may hold in some cases, there is evidence that it does not apply to many of the reactions so far studied (Bowden 1928).

(5) *The dependence of η_c on I.* In general the relation between the overpotential and the current density for concentration overpotential will be different from that observed for activation overpotential. One exception to this statement has been mentioned (p. 214 and fig. 2).

(6) *The rate of growth of overpotential.* In a similar manner, the potential-time relation for concentration overpotential, illustrated in fig. 3, is generally quite different from that observed in systems in which activation overpotential is of primary importance. As Bowden and Rideal (1928), and others, have shown, activation overpotential increases linearly with the time, and the electrode has a "capacity" of the order 20 μF/cm.^{2}. For concentration overpotential, the potential-time relation is not linear, and the rate of growth is generally very much slower, corresponding to an apparent "capacity" of several thousand microfarads per sq. cm.

SUMMARY

We can distinguish three main causes of irreversibility in electrode reactions. These are:

(i) "Concentration overpotential."
(ii) "Activation overpotential."
(iii) "Resistance overpotential."

A general theory of concentration overpotential is developed, and this shows how the overpotential will depend on the current density, concentration and temperature. An expression for the rate of growth of concentration overpotential after switching on the current is also given.

It is shown that the resistance of the diffusion layer may in some cases attain high values, and cause a resistive potential drop at least as great as the concentration overpotential. Finally, the principal differences between concentration overpotential and activation overpotential are described.

PART II. THE ELECTRODEPOSITION OF OXYGEN FROM FUSED SODIUM HYDROXIDE

The electrolytic deposition of the hydroxyl ion from aqueous alkali solutions has been studied by Hoar (1933), and by Bowden and Keenan (unpublished work). These authors find that the deposition, and the reverse process (electro-solution of oxygen) have high overpotentials at current densities as low as 10^{-6} amp./sq. cm.; in both cases the overpotential varies linearly with the logarithm of the current density. In this and other respects, the overpotential shows the characteristics of "activation" overpotential in the sense defined in Part I.

In aqueous solution, however, the ions may be hydrated and the electrode process is necessarily complicated by the presence of water. It is therefore of interest to investigate the corresponding process in a system such as fused sodium hydroxide, which consists of "bare" sodium and hydroxyl ions. The experimental difficulties of work with fused alkalis are considerable, and, probably for this reason, no detailed study of electrode processes in such a system has previously been made. Haber (1906) and his collaborators measured the reversible e.m.f. of an oxygen-hydrogen cell in fused caustic soda, in equilibrium with water vapour at a known partial pressure, and showed that it agreed with the value calculated thermodynamically. This shows that an oxygen electrode takes up its correct reversible potential under these conditions. Le Blanc and Brode (1902), Sacher (1901) and others have studied the electrolysis of fused sodium hydroxide. Their results indicate that the deposition of oxygen occurs irreversibly under these conditions, but they do not give us any quantitative information.

The present experiments show that the deposition of oxygen from fused sodium hydroxide is accompanied by a considerable overpotential and that this overpotential varies in an orderly manner with the current density.

The detailed study of the reaction makes it clear that "activation overpotential" is negligibly small throughout the large range of current densities used. The overpotential observed is actually "concentration overpotential" of a rather curious kind. This is in complete contrast with the conclusions to be drawn from investigations of the corresponding reaction in aqueous solutions. The results of the present work support the general theory of concentration overpotential given in Part I.

APPARATUS AND EXPERIMENTAL METHOD

Since glass, quartz, porcelain, etc., are rapidly attacked by fused caustic alkalis, substances of this type could not be used in the construction of the cell. Nickel was found to be suitable for this purpose. Under the conditions of these experiments nickel rapidly becomes covered with a black oxide film, but further oxidation is very slow.

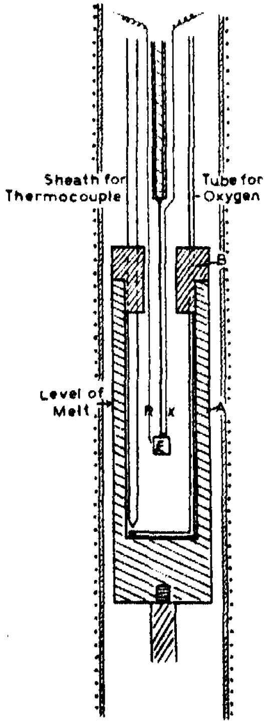


FIG. 4. The cell.

The fused alkali was contained in the nickel pot, *A* (fig. 4). The anode *E* on which the measurements were made, consisted of a rectangular piece of nickel or platinum sheet mounted on thick wire of the same material. The reference electrode which is represented diagrammatically by the line *R* in fig. 4, was a fine wire, either nickel or platinum. This acted as a reversible oxygen electrode. The nickel pot *A* itself served as cathode. The whole cell was supported inside a vertical tube of transparent quartz. The quartz tube was heated externally by an electrical heating element, so that the cell could be raised to any desired temperature. The ends of the tube were closed, to exclude atmospheric carbon dioxide and water vapour. The polarizing current was taken from a high voltage supply, and could be varied over a large range with the aid of a potential divider and resistances. The overpotential is given directly by the potential difference between the anode, *E*, and the reversible reference electrode. In the experiments on the rate of growth of overpotential an Einthoven galvanometer and moving paper camera, as described by Bowden and Rideal (1928) were used to follow the rapid changes of potential.

Pure dry oxygen was continuously bubbled through the melt when the cell was in operation. Before entering the cell the oxygen stream could if desired, be diverted through a bulb containing water. In this way small quantities of water vapour could be transferred to the melt.

Pure "sodium hydroxide from sodium" was used. When added to the cell it contained a trace of carbonate and up to 1 % of water. Most of the latter evaporated on heating for a few hours at 350° to 400° in the stream of dry oxygen.

EXPERIMENTAL RESULTS

(*a*) *The variation of electrode potential with current density*

Immediately after the beginning of an experiment, the values of the overpotential observed are low and erratic. After a few hours, however, they rise and attain a fairly constant value. This change is no doubt due to the loss of water mentioned above. The behaviour at high anodic current densities, after the steady state is reached, is shown in fig. 5, in which the overpotential, η, is plotted against the logarithm of the current density, I. In these experiments the temperature ranged from 318 to 340° C, and the melt was in all cases stirred vigorously.

At the higher current densities the resistive potential drop in the melt cannot be neglected. This potential drop was measured directly by a movable reference electrode. The observed potential was found to vary linearly with the distance between the electrodes. From these observations it is possible to calculate the specific conductivity of the melt. The value found was 2·5 mho/cm. cube at 320° C. Arndt and Ploetz (1926) give 2·25 mho/cm. cube. This agreement is as good as can be expected. Since the distance (about 0·1 cm.) between the reference electrode and the anode is known, it is possible to use these results to apply a correction for the resistive potential drop in the electrolyte. The values of η for nickel and bright platinum electrodes plotted in fig. 5 are corrected in this way. The correction is negligible at the lower current densities and amounts to about 30 mV at the highest current density employed.

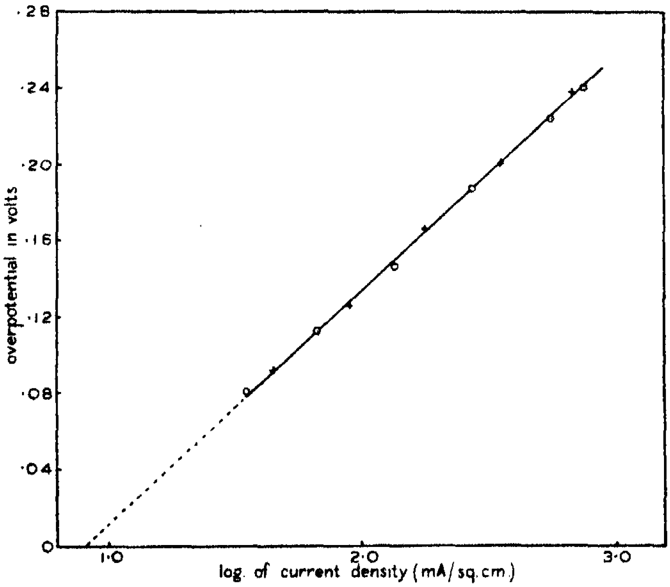


Fig. 5. $\eta - \log I$ curves for nickel and platinum anodes. ⊙ Ni anode (corrected for resistance of melt). × Pt anode (corrected for resistance of melt).

It is evident that there is a linear relation between the corrected values of the overpotential, η, and $\log I$ for overpotentials higher than about 0·100 V. This can be represented by the equation

$$\eta = b \log \frac{I}{A}. \tag{1}$$

As will be seen from fig. 5, the overpotentials on bright platinum and nickel surfaces are the same to within the experimental error. The values of A observed range from 4×10^{-3} to 9×10^{-3} amp./sq. cm., and b from 0·110 to 0·127 V. The mean value of b found is 0·121 V. It is necessary to add that the values of A quoted do not have any absolute significance, since it will be shown that A depends on the intensity of stirring of the melt. The values given here and elsewhere refer to a constant rate of flow of oxygen through the cell, and therefore to constant stirring conditions.

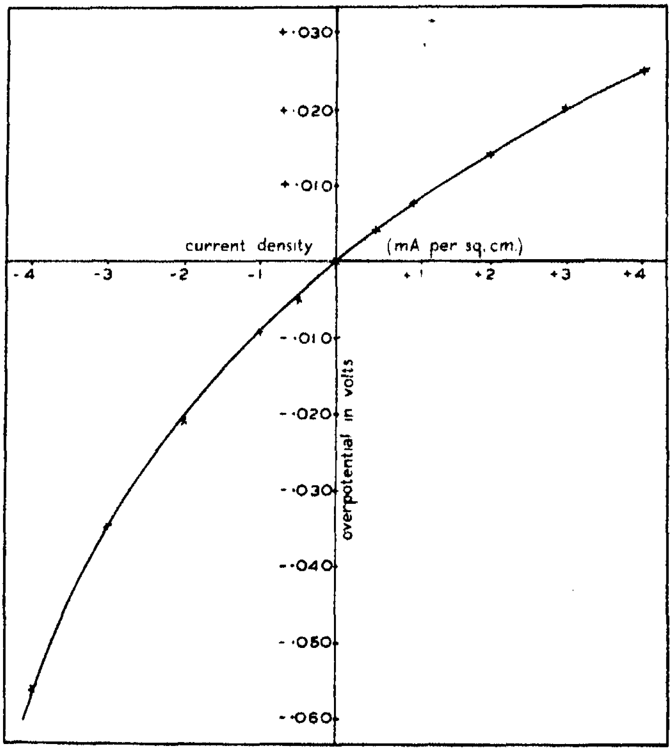


FIG. 6. Relation between η and I at lower current densities (Ni anode).

At low current densities the logarithmic relation no longer holds. The behaviour under these conditions is shown in fig. 6, in which the overpotential, η, is plotted against the current density, I. Positive values of I and η refer to the anodic reaction, i.e. the discharge of OH′ ions with the formation of oxygen and water; negative values to the reverse reaction, i.e. the formation of OH′ ions from water and oxygen gas. At these low current densities the resistance correction is negligible.

A few observations have been made using platinized platinum anodes. Within the limits of the experiments, the value of the overpotential is the same as on bright platinum or nickel.

(*b*) *The influence of stirring*

The melt in the cell was continuously stirred by the stream of oxygen flowing through it. It was found that increasing the stirring by increasing the rate of flow of oxygen always reduced the overpotential. This effect was particularly marked at low rates of flow. The rate of flow could be controlled by altering the pressure drop across a capillary tube, and was generally kept at a predetermined value throughout each experiment.

Expressing the rate of flow in arbitrary units equal to the pressure drop across the gauge, we find that the overpotential increases by about 25 % when the rate is reduced from 4 to 0·5 unit. This corresponds to a decrease of about 50 % in A. The numerical values of A, etc., quoted, refer to a flow rate of 5 units.

(*c*) *The effect of temperature*

Owing to irreversible changes in the melt (probably due to loss of water), which occur when the temperature is raised much above 350° C, it is difficult to obtain accurate values of the temperature coefficient of the overpotential. A number of determinations showed, however, that it did not differ much from $-0{\cdot}3$ mV/° C.

(*d*) *The rate of growth of overpotential*

These experiments were carried out with a moving paper camera and an Einthoven galvanometer. Fig. 7*a* (Plate 9) shows a typical potential-time curve for the growth of the overpotential on a nickel surface in fused sodium hydroxide. The current density used is 3×10^{-2} amp./sq. cm. The Einthoven galvanometer shows that there are slight oscillations in the final "steady" value of η. These can probably be ascribed to the passage of separate bubbles of oxygen past the electrode surface. They are, of course, too rapid to affect an ordinary galvanometer.

For comparison, figs. 7*b* and 7*c* (Plate 9) show the growth of oxygen overpotential in *aqueous solutions* (activation overpotential).

(*e*) *The effect of the addition of water*

The results so far described refer to specimens of sodium hydroxide containing only a small residual trace of water. Larger quantities (of the order of 1 %) could be added by evaporation from the flask in the oxygen train. The effect of adding water vapour in this way is very striking. For example, in one experiment, at a current density of 0·020 amp./sq. cm., the overpotential fell from 0·045 to 0·007 V within a few seconds of diverting the oxygen stream through the flask. A sudden fall of this kind is perfectly general, and shows that a very small amount of water added to the melt has a pronounced effect on the overpotential. Curve *B*, fig. 8, shows the overpotential at constant current density, with various amounts of water added.

(*f*) *Melts containing sodium oxide*

Sodium oxide was prepared in situ by adding a small quantity of metallic sodium to the melt, and oxidizing it in a stream of oxygen. In this way we obtain a melt containing about 1 % of the oxide. This concentration could be readily diminished by the addition of small amounts of water in the manner described above. The oxide is thereby converted into sodium hydroxide. As is clear from figs. 8 and 9, the overpotential in oxide containing melts is much less than that in untreated hydroxide. In this respect sodium oxide has an action similar to that of water.

The data plotted in fig. 8 were obtained by adding water to a melt initially containing oxide. As the concentration of this falls the overpotential rises, reaching a sharp maximum when the composition corresponds to pure NaOH. Further addition of water causes the overpotential to decrease again, in the manner already described.

DISCUSSION

(*a*) *The nature of the electrode reaction*

Owing to the occurrence of the reaction

$$2OH' = H_2O + O'',$$

we must expect to find a small concentration of water and of O″ ions in a pure NaOH melt. Two lines of evidence show, however, that the degree of dissociation of the OH′ ion in this manner must be extremely small: (*a*) pure NaOH does not lose water, even at very high temperatures, and (*b*) sodium metal does not react with pure sodium hydroxide at 300–400° C (Le Blanc and Brode 1902). These facts indicate that the concentration of water is vanishingly small in pure fused sodium hydroxide; the concentration of O″ must be equally small. In general, sodium hydroxide at room temperature contains an additional trace of water. Under these conditions the concentration of O″ is even further reduced, and it is safe to assume that is is small compared with that of the water. We are therefore justified in supposing that the O″ ions from the bulk of the melt do not take part to

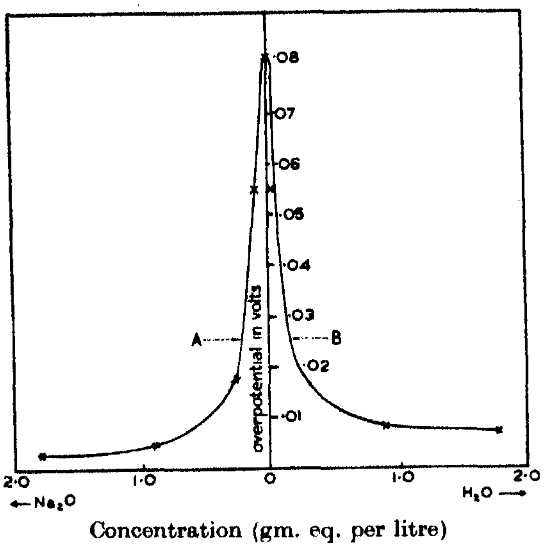


FIG. 8. Fall of overpotential on addition of water or sodium oxide.

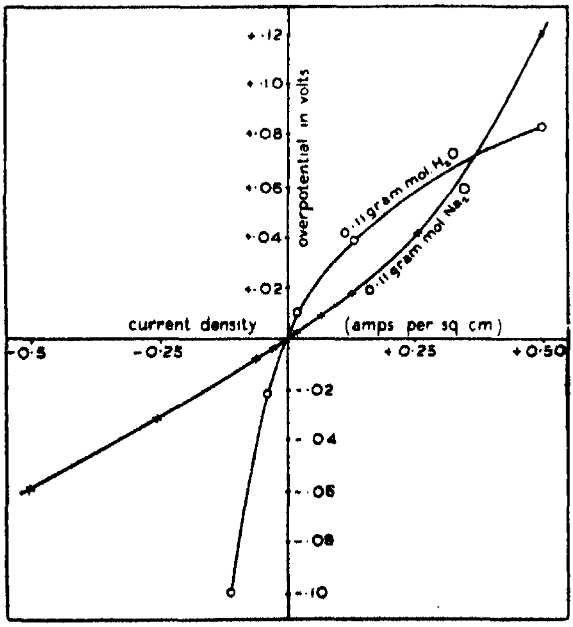


FIG. 9. η-I curves in melts containing water and sodium oxide.

a significant extent in the electrode reaction in such a system. As far as the anodic reaction is concerned, these ions could not be brought up to the surface sufficiently rapidly. In the cathodic reaction, the solution of oxygen as O'' ions would lead to an accumulation of these ions near the electrode. This can only occur at potentials much more negative than the reversible oxygen potential in this system. In melts containing a trace of water, therefore, the electrode reaction may be taken to be the discharge of OH' ions when the electrode is anodically polarized, and the formation of these ions, from oxygen and water, when the electrode is cathodically polarized.

Some of the experimental evidence already presented, e.g. the influence of stirring, and the fact that the overpotential is independent of the nature of the anode surface, suggests strongly that the overpotential is due to concentration changes near the electrode. The following treatment is based on this view, and leads to results which provide further confirmation of its correctness.

(b) *The relation between η and I*

The equations for concentration overpotential developed in Part I can easily be applied in the present case. The anode reaction can be written

$$-4OH' + 4\epsilon + O_2 + 2H_2O = 0.$$

In the system considered the activity of the OH' ion cannot vary appreciably; any concentration overpotential must be due to the oxygen and/or the water. As these are not ionized, transport by migration and diffusion potentials do not have to be considered. The oxygen readily escapes in the stream of gas passing through the cell, and on this account its concentration at the electrode will probably not exceed the saturation value by very much. To a first approximation, we may therefore neglect the effect of this product, and consider only that of the water.

In the theory developed in Part I we assumed that the activity of the substances taking part in the reaction was equal to their concentration. The activity of water dissolved in fused sodium hydroxide can be determined from measurements of the vapour pressure of water in the melt. The results of several investigations are given in *International Critical Tables* (**3**, 370). Fig. 10, in which the logarithm of the vapour pressure p is plotted against the logarithm of the concentration of water, c, is constructed from the data given there. In this range the curve has a slope varying from 1·5 to 3·0, and it is thus evident that the vapour pressure (or activity) is by no means directly proportional to the concentration.

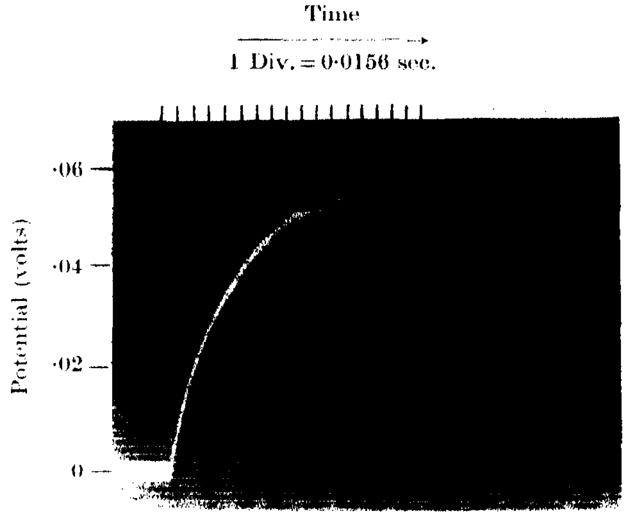


Fig. 7*a*. Ni anode in fused NaOH. c.d.: $3\cdot0 \times 10^{-2}$ amp./sq. cm.

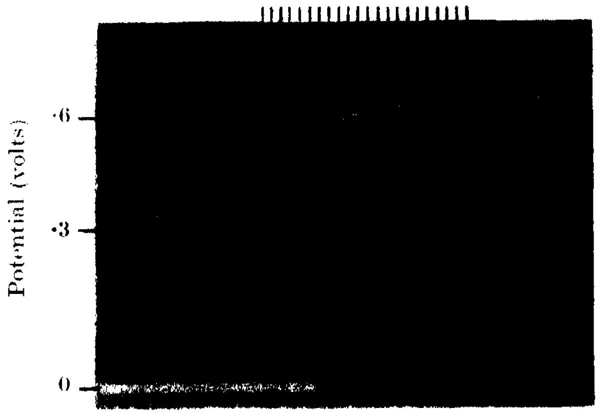


Fig. 7*b*. Pt anode in N/10 NaOH soln. c.d.: $2\cdot5 \times 10^{-2}$ amp./sq. cm.

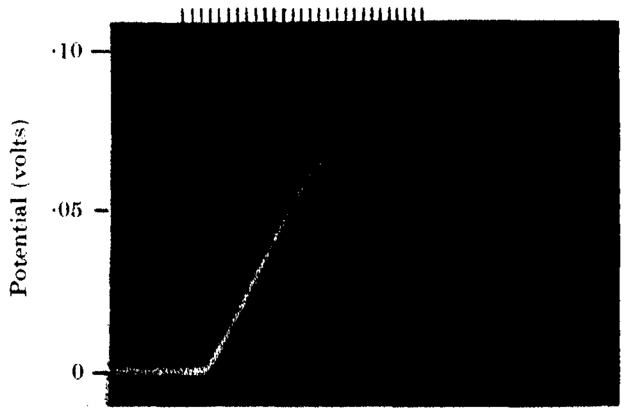


Fig. 7*c*. Pt anode in N/10 NaOH soln. c.d.: $7\cdot7 \times 10^{-6}$ amp./sq. cm.

If we consider two arbitrary systems, represented by the points A and B in fig. 10, and write

$$\frac{p_A}{p_B} = \left(\frac{c_A}{c_B}\right)^{\lambda}, \tag{2}$$

we find that the index λ lies between 1·5 and 3·0 in the region covered by the curve. If we suppose that A corresponds to the concentration of water at the anode and B to that in the bulk of the melt, we have

$$\eta = \frac{RT}{2F}\ln\frac{p_A}{p_B} = \frac{RT}{2F}\ln\left(\frac{c_e}{c_0}\right)^{\lambda}. \tag{3}$$

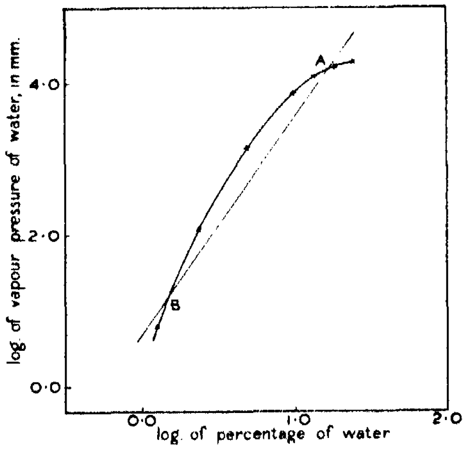


FIG. 10. The vapour pressure of water over fused sodium hydroxide.

Introducing the previous value for c_e/c_0 we get

$$\begin{aligned} \eta &= 0{\cdot}084\log\left(1-\frac{I}{I_0}\right)\text{ volts,}\quad \text{if } \lambda = 1{\cdot}5 \\ &= 0{\cdot}177\log\left(1-\frac{I}{I_0}\right)\text{ volts,}\quad \text{if } \lambda = 3{\cdot}0. \end{aligned} \tag{4}$$

If $I \gg -I_0$ equation (4) reduces to

$$\eta = 0{\cdot}059\lambda\log\frac{I}{-I_0}. \tag{5}$$

(I_0 is a negative quantity; see equation (10) of Part I.) This relation is of the same form as that observed at high current densities (equation (1)). The constant A in (1) is to be identified with $-I_0$. The observed value of the quantity b in (1), (0·121 V) lies between the extreme values of the coefficient in (4) (0·084 and 0·177 V).

The curve in fig. 6, which refers to lower current densities, represents the relation

$$\eta = b \log\left(1 - \frac{I}{I_0}\right), \tag{6}$$

with $b = 0{\cdot}110$ V, $I_0 = -5{\cdot}8 \times 10^{-3}$ amp./sq. cm. Within the experimental error these values are the same as those found from the linear $\eta - \log I$ curve at higher anodic current densities. It will be seen that the experimental points (×) in fig. 6 fit the curve very well. It is thus clear that the overpotential obeys a relation of the same type as the theoretical equation (4) throughout the range of current density available to investigation. Furthermore, the value of the quantity b is of the order we should expect on theoretical grounds, provided the difference between the activity and concentration of the dissolved water is taken into account.

(c) *The value of I_0, and the thickness of the diffusion layer*

Reference to equation (10) of Part I shows that

$$I_0 = -\frac{2Fkc_0}{\delta}. \tag{7}$$

We have little knowledge of the magnitude of c_0; it is, however, possible to make a rough estimate of the values of k and δ from independent sources. A few measurements of diffusion coefficients in fused salts have been made (see Landolt-Börnstein, *Tabellen*, Ergänzungsbd. **2**, 199). Values of the order $4{\cdot}5 \times 10^{-5}$ cm.^{2}/sec. are observed under conditions similar to those of the present experiments. Knowing k, we can calculate the experimental value of δ from the time constant, μ. From the potential-time curves of fig. 7 we find approximately

$$\frac{\pi^2 k}{4\delta^2} = \mu = 10\,\text{sec.}^{-1}.$$

Taking $k = 4{\cdot}5 \times 10^{-5}$ cm.^{2}/sec., we obtain $\delta = 0{\cdot}003$ cm.

The thickness of the diffusion layer depends on the hydrodynamical properties of the liquid, but the available evidence indicates that it is not very sensitive to changes in these. Since the viscosity and density of fused sodium hydroxide at 320° C only differ by a factor of 2 or 3 from the corresponding quantities for aqueous solutions at room temperature, we may reasonably expect that, under similar conditions of stirring, δ will be of the same order in both liquids. In fact, many workers have observed values of between 0·001 and 0·005 cm. in vigorously stirred aqueous solutions, which agrees with the above experimental value (0·003 cm.) for fused sodium hydroxide.

Taking $\delta = 0{\cdot}003$ cm., and using the experimental value of I_0 $(-6\times 10^{-3}$ amp./sq. cm.), equation (7) shows that:

$$\begin{aligned} c_0 &= 2\times 10^{-6}\,\text{g.-mol. H}_2\text{O per c.c.} \\ &= 2\times 10^{-3}\,\text{g.-mol. H}_2\text{O per litre.} \end{aligned}$$

At the maximum overpotential measured in these experiments, the concentration of water at the electrode must be about 100 c_0, i.e. 2×10^{-1} g.-mol./l., or nearly 0·19 % by weight.

It is evident that, if all the water liberated at the anode were to evaporate in the bubbles of gas arising from the electrode, its partial pressure would be $\frac{2}{3}$ atm. Such a pressure is only reached, however, when the concentration of water in the melt is of the order 5 % (*International Critical Tables. loc. cit.*). As the maximum concentration at the electrode in these experiments is found to be 0·19 %, it is clear that very little water will be lost from the anode region by evaporation. In confirmation of this conclusion, Le Blanc and Brode (1902) have found experimentally that the gases liberated at the anode do not contain an appreciable quantity of water. We are thus justified in assuming that the water is removed from the anode by diffusion and convection. On the addition of 1 % of water to the melt, c_0 is increased by 100 or perhaps 1000 times. Equation (7) shows that there will be a similar increase in I_0. We can thus account in a very simple manner for the pronounced fall of overpotential which occurs when small quantities of water are added to the melt.

In the melts containing sodium oxide the electrode reaction must be of a different nature. In these the concentration of O″, or of Na_2O, is much greater than that of water. Under these circumstances water is not of much importance in the electrode reaction, and need not be considered further. For simplicity we assume that the anodic reaction is the discharge of O″ ions. The argument is not affected if the substances taking part in the reaction are in reality undissociated Na_2O molecules, or conceivably Na_2O_2. Taking the electrode reaction to be

$$2O'' = O_2 + 4\epsilon,$$

and using equation (13) of Part I, we have

$$\eta = -\frac{RT}{2F}\ln\left(1-\frac{I}{I_0}\right), \tag{8}$$

where the constant I_0 now refers to the O″ ion. This I_0 is a positive quantity; that for water in the water-containing melts is negative (see equation (7)). Owing to this difference in sign, we should expect that the $\eta - I$ curves in the melts containing oxide will be concave towards the positive axis; in the ordinary, water-containing melts they are convex towards this axis. This prediction is confirmed by the experimental curves (fig. 9).

Independent confirmation of the view that the overpotential in the water-containing melts is due to local changes in the concentration of water is provided by the low value of the temperature coefficient, the effect of stirring, and the rate of growth of the overpotential. The observed value of the temperature coefficient is approximately $-0{\cdot}3$ mV/degree. This is about the value we should expect for a concentration overpotential. It is much smaller than that characteristic of activation overpotential, which is usually of the order -2 to -3 mV/degree (Bowden 1929). The marked reduction in the overpotential which is brought about by increased stirring is highly characteristic of concentration overpotential, but is difficult to explain on any other basis.

The rate of growth of activation overpotential has been investigated by Bowden and Rideal (1928), Bowden and Grew (unpublished work) and others. It is found that the potential change is directly proportional to the time, or quantity of electricity passed, over a large range. The "capacity" of the electrode is of the order 20 μF/sq. cm. for mercury cathodes. Figs. 7*b* and 7*c* show Einthoven photographs of the growth of oxygen overpotential (activation overpotential) on a platinum anode in N/10 NaOH solution at room temperature. They are included for comparison with the growth curve in fused NaOH (fig. 7*a*). In fig. 7*b* the current density employed is practically the same as in fig. 7*a*. It is clear that in the aqueous solution the change of potential is so rapid that it cannot be followed. Fig. 7*c* refers to a lower current density ($7{\cdot}7 \times 10^{-6}$ amp./sq. cm.). It will be seen that the trace is substantially linear over most of the increment in electrode potential. From the slope the capacity is found to be 24 μF/sq. cm. In fused sodium hydroxide (fig. 7*a*) the potential time curve is not linear, and the mean "capacity" of the electrode (up to half the final potential) is about 40,000 μF/sq. cm. It should further be noted that the curve in fig. 7*a* is of the same type as the theoretical curves for concentration overpotential (fig. 3 of Part I).

Finally, it should be noted that the independence of the overpotential on the nature of the electrode surface is entirely to be expected if the overpotential is concentration overpotential, but would be difficult to explain if it were of any other type.

We wish to acknowledge a grant for apparatus from the Chemical Society, and one of us (J.N.A.) is indebted to the Department of Scientific and Industrial Research for a Maintenance Grant and to Sidney Sussex College for a Research Studentship.

SUMMARY

Measurements of the oxygen overpotential on nickel and platinum electrodes in fused sodium hydroxide have been made. The current densities used ranged from very low values up to 1 amp./sq. cm. At the highest current density the overpotential is of the order 0·2 V.

Although in some respects the observations might suggest that the overpotential is of the same kind as that observed in the deposition of OH′ ions in aqueous solutions, the evidence as a whole shows that this is not so. The overpotential in fused NaOH is due to the change in the *concentration* of *water* in the melt in the immediate vicinity of the electrode. A small trace of water is normally present in the fused hydroxide, and more water is produced by the anode reaction.

This view is supported by many different lines of evidence, including:

(*a*) The influence of stirring the electrolyte.
(*b*) The effect of adding small quantities of water.
(*c*) The rate of increase of the overpotential on switching on the current.
(*d*) The nature of the relation between overpotential and current density.
(*e*) The low value of the temperature coefficient.
(*f*) The fact that the overpotential is independent of the nature of the electrode.

The general behaviour is consistent with the theory of concentration overpotential developed in Part I.

REFERENCES

Arndt and Ploetz 1926 *Z. phys. Chem.* **121**, 437.
Bowden 1928 *Trans. Faraday Soc.* **24**, 473.
— 1929 *Proc. Roy. Soc.* A, **126**, 107.
Bowden and Rideal 1928 *Proc. Roy. Soc.* A, **120**, 59.
Brunner 1904 *Z. phys. Chem.* **47**, 56.
Glasstone and Hickling 1935 "Electrolytic Oxidation and Reduction", ch. III.
Haber 1906 *Z. anorg. Chem.* **51**, 245, 289, 356.
Haskell 1908 *Phys. Rev.* **27**, 145.
Hoar 1933 *Proc. Roy. Soc.* A, **142**, 628.
Le Blanc and Brode 1902 *Z. Elektrochem.* **8**, 697, 717.

Nernst 1888 *Z. phys. Chem.* **2**, 612.

— 1904 *Z. phys. Chem.* **47**, 52.

Noyes and Whitney 1897 *Z. phys. Chem.* **23**, 689.

Roseburgh and Lash Miller 1910 *J. Phys. Chem.* **14**, 816.

Sacher 1901 *Z. anorg. Chem.* **28**, 385.

Sand 1901 *Phil. Mag.* **1**, 45.

Volmer and Wick 1935 *Z. phys. Chem.* A, **172**, 429.

Wirtz 1938 *Z. Elektrochem.* **44**, 303.

The effect of solvents on the continuous absorption spectrum of bromine

By R. G. Aickin, N. S. Bayliss and A. L. G. Rees
Chemistry Department, University of Melbourne

(*Communicated by F. G. Donnan, F.R.S.—Received* 26 *July* 1938)

Introduction

The following reasons prompted us to select bromine as the subject for a study of the effect of the solvent on the absorption spectrum of a dissolved substance:

(1) Bromine has a simple molecule, and its absorption spectrum in the gaseous state has been interpreted fairly satisfactorily (Acton, Aickin and Bayliss 1936; Aickin and Bayliss 1938; Darbyshire 1937; Mulliken 1936).

(2) It has continuous absorption in a convenient spectral region (5000 to below 2200 A), where many common solvents are transparent. There are advantages in working with a continuous spectrum, since it is possible to compare absorption coefficients under varying conditions without the difficulties that are associated with the blurring out of band structure (Rabinowitch and Wood 1936).

(3) The continuous absorption of bromine is complex, being composed of transitions to several excited states (Acton, Aickin and Bayliss 1936; Aickin and Bayliss 1938). This makes possible a direct comparison of the effect of a solvent on different electronic transitions.

It was fortunate that while this work was in progress, the work of Porret (1937) was published, enabling a further comparison to be made with liquid bromine.

This paper presents an account of the effect on the absorption spectrum of bromine, at concentrations between 0·00015 and 0·0010 mol./l., of the following solvents: 2N sulphuric acid, cyclohexane, chloroform, carbon tetrachloride, benzene, chlorobenzene and toluene. The main effect in the case of 2N sulphuric acid will be that of the water; an acid solution of this particular concentration was chosen in order to suppress the hydrolysis of bromine to less than one part in 1000 (Liebhafsky 1934). Other solvents were tested, such as carbon bisulphide, several alcohols and bromoform. Carbon bisulphide was rejected because it absorbs light strongly throughout a considerable part of our chosen spectral region, while in the case of the alcohols, even tertiary butyl alcohol reacted too rapidly with bromine. Bromoform gave trouble owing to a fairly rapid photo-decomposition that resulted in the liberation of bromine, and it also was discarded.

EXPERIMENTAL

Our materials were all carefully purified. The refractive indices recorded below were determined by a Zeiss-Pulfrich refractometer, and are compared with the values given by Weissberger and Proskauer (1935), denoted below by W.P.

Bromine. The sample was A.R. bromine that had been previously distilled over a mixture of pure calcium bromide and calcium bromate by Professor E. J. Hartung, to whom we are indebted for the sample. It was subsequently distilled *in vacuo* several times, the head and tail fractions being discarded. The tail fraction contained less than one part of iodine in 5 million of bromine.

Cyclohexane. Since the most likely impurities are unsaturated hydrocarbons, particularly cyclohexene, which have a much lower freezing point than cyclohexane, the product was purified by fractional freezing, followed by drying over fused calcium chloride. $n = 1{\cdot}42606$ (20° C) (W.P. 1·42636).

Chloroform. B.D.H. A.R. chloroform was repeatedly washed with distilled water and then dried over calcium chloride. It was then fractionated, collected over calcium chloride, and kept in the dark until used. According to Twyman and Allsopp (1934), chloroform prepared in this way becomes useless for spectrographic work after 12 hr. We therefore distilled this solvent on the day on which it was to be used. $n = 1{\cdot}44808$ (15° C) (W.P. 1·44858).

Carbon tetrachloride. Merck's guaranteed reagent was fractionated, and stored, over potassium hydroxide. $n = 1{\cdot}45966$ (20° C) (W.P. 1·46040).

Benzene. Merck's reagent was guaranteed free from all sulphur compounds. It was dried over freshly cut sodium for several days before use. $n = 1{\cdot}50388$ (15° C) (W.P. 1·50440).

Chlorobenzene. A good commercial sample of chlorobenzene was fractionated and dried over calcium chloride. The sample had a constant boiling point of 132° C.

Toluene. In order to remove sulphur compounds from the commercial sample, it was shaken with two portions of concentrated sulphuric acid for 30 hr. each, followed by a shaking with distilled water for 12 hr. After repeated washing with distilled water, it was dried over calcium chloride, fractionated, and the product dried over sodium. It did not appreciably decolorize bromine during the time of an exposure. $n = 1{\cdot}49833$ (15° C) (W.P. 1·49985).

METHOD

Our method of photographic spectrophotometry consisted, in outline, of passing a beam of suitable continuous radiation through a cell containing the liquid under examination, and then on to the slit of the Hilger all-metal medium spectrograph. On each plate (10 × 4 in.) we obtained about eight exposures, all with the same time of exposure. One of them was the absorption spectrum of the bromine solution. For the others, the cell was emptied, refilled with the pure solvent, and the light beam was weakened by known amounts by interposing calibrated wire-gauze screens, which were mechanically shaken eccentrically in a plane normal to the path of the light. In addition to the above exposures, a spectrum of a brass or iron arc was photographed at each side of each plate, in order to provide wavelength calibrations. After development, photometer traces were taken across each plate between corresponding lines of the arc spectra.

The transmission of the bromine solution at each chosen wave-length was thus compared directly with the transmissions of the calibrated screens, and the actual values for the solution were determined by graphical interpolation between the known screen values. After the concentration of each solution was obtained, by shaking with aqueous potassium iodide and titration against standard sodium thiosulphate, the absorption coefficients were calculated from the relation $\log_{10}(I_0/I) = \epsilon c d$, where ϵ is the absorption coefficient, c is the concentration of bromine in mol./l., d is the length of the absorbing column in cm., and I_0 and I are the incident and transmitted intensities respectively. In different experiments, our bromine concentrations ranged from 0·00015 to 0·0010 mol./l.

The absorption due to the solvent itself was corrected for by taking the calibration spectra through the pure solvent, contained in the cell that had been used to contain the solution. The cell holder was so designed that it was easy to remove the cell and replace it accurately in the same position with respect to the optical path.

As sources of continuous radiation we used a 30-watt car headlamp or a water-cooled hydrogen tube, according to the spectral region. In each case, the applied voltage was hand regulated for constancy during a run. The absorption cell was of silica, had plane windows, and was 5 cm. long (Hilger "Presil", type D). In the experiments that were performed to fix the positions of the ultra-violet maxima, and in which we did not determine absorption coefficients, the solutions were contained in an all-silica Baly cell by Hilger. Exposure times were controlled electrically by means of a synchronous electric clock, and exposures of any integral multiple of 30 sec. could be repeated with an error not exceeding 0·2 sec. Both quartz and glass optical systems were used with the spectrograph according to the region of the spectrum. Previous work had shown us that stray light in the spectrograph introduced errors owing to the reception by "unexposed" portions of the plate of radiation sufficient partly to overcome the inertia of the emulsion. In the present experiments, we guarded against this error by constructing a diaphragm that fitted into the camera of the spectrograph, and which shielded all but the strip of plate, about 0·5 cm. wide, on which the spectrum was being photographed. The calibration of the wire-gauze screens, and the construction of the photoelectric photometer, have been described previously (Acton, Aickin and Bayliss 1936).

Our technique included the precautions that are necessary in photographic photometry (Harrison 1929). Blackened diaphragms placed at intervals along the optical path minimized the amount of stray light reaching the slit of the spectrograph. Kodak process, Universal and Ordinary plates were used. In no case did the emulsion contain a dye which after development might provide a background of uneven transmissivity. A metol-hydroquinone developer was used, and the plates were brushed lightly to ensure even development. After hardening and washing, they were swabbed with distilled water and slowly dried.

RESULTS

Our absorption coefficients, defined as above, are given in Table I. The absorption curves of bromine in the various solvents are shown in figs. 1–3. In each case, a comparison is drawn with the absorption of gaseous bromine, which is represented by the broken curves derived from the results of

TABLE I

Wave-length Å	Absorption coefficients of bromine dissolved in							Wave no. cm.$^{-1}$
	Water	Cyclo-hexane	Chloro-form	Carbon tetra-chloride	Benzene	Chloro-benzene	Toluene	
5292	15·8	38·5	34·8	42·8	40·9	77·1	75·7	18889
5250	19·5	34·2	37·1	48·6	33·8	79·9	78·0	19042
5220	22·6	39·9	46·4	52·5	69·4	80·7	90·0	19152
5153	25·1	44·7	55·7	60·8	78·9	91·0	105·9	19400
5106	29·4	52·0	62·2	72·8	82·5	97·4	111	19581
5076	31·2	56·1	—	74·0	85·3	105	120	19695
5016	37·8	61·5	69·1	82·0	96·9	114	131	19928
4958	40·3	67·6	80·6	92·5	108	122	146	20165
4919	44·8	72·5	87·2	98·0	122	134	151	20324
4865	50·5	83·5	94·2	103	131	141	161	20549
4810	55·9	86·4	100	108	142	150	169	20782
4768	62·0	104	108	115	153	159	183	20971
4722	67·2	110	111	119	166	176	196	21171
4651	78·3	—	119	127	190	184	215	21494
4587	86·1	127	128	133	209	199	228	21795
4540	93·7	137	133	139	218	206	234	22022
4480	103	150	141	148	232	216	255	22314
4415	111	164	154	156	248	229	272	22643
4378	116	178	162	163	255	231	285	22834
4326	122	194	171	174	265	244	301	23111
4275	128	—	180	176	278	250	326	23385
4226	133	212	189	181	299	258	336	23652
4178	136	—	192	182	—	258	351	23930
4144	139	—	203	193	309	261	358	24125
4104	—	—	—	—	—	263	—	24358
4063	147	207	206	203	302	262	—	24603
4023	153	201	196	200	294	—	354	24852
3969	156	189	184	185	280	260	353	25186
3930	164	158	172	169	—	246	364	25436
3862	157	—	148	152	260	227	390	25887
3801	152	138	125	129	239	208	448	26301
3734	144	86·7	100	105	227	184	—	26767
3684	130	—	90·6	88·5	231	200	708	27142
3640	120	71·2	78·5	70·9	246	213	859	27468
3602	111	63·6	71·7	58·6	292	222	910	27754
3527	90·3	59·1	49·5	—	328	312	—	28341
3476	78·4	37·7	44·9	42·5	633	418	—	28765
3441	70·0	38·6	40·5	38·5	739	509	—	29053
3393	54·3	34·1	44·2	37·9	1048	677	—	29467
3338	41·1	51·4	38·6	32·1	1433	897	—	29951
3293	30·0	61·7	36·2	33·5	1690	—	—	30350
3274	24·5	62·2	—	33·5	—	—	—	30535
3248	21·8	85·7	—	32·8	—	—	—	30783
3208	18·6	102	52·7	33·9	—	—	—	31161
3175	19·5	113	65·2	37·6	—	—	—	31486
3147	20·1	130	78·5	40·3	—	—	—	31769
3126	16·0	146	93·5	46·2	—	—	—	31979
3094	19·6	158	115	39·9	—	—	—	32311
3063	23·0	182	143	39·7	—	—	—	32634
3036	26·9	206	168	43·0	—	—	—	33927
2997	33·3	233	201	43·4	—	—	—	33353
2961	40·1	277	—	49·7	—	—	—	33760
2937	44·1	296	—	54·9	—	—	—	34040
2914	52·1	322	—	54·7	—	—	—	34312
2883	61·5	363	—	51·5	—	—	—	34676
2863	72·6	457	—	56·9	—	—	—	34914
2824	85·1	(418)	—	58·2	—	—	—	35396
2801	94·4	—	—	59·4	—	—	—	35691
2780	105	—	—	67·8	—	—	—	35963

Acton, Aickin and Bayliss (1936) below 30,000 cm.$^{-1}$. Above this frequency, the absorption of gaseous bromine is shown diagrammatically, since it is too weak to be plotted accurately on our chosen scale of ϵ (Aickin and Bayliss 1938). In fig. 3 the upper broken curve is a relevant part of the absorption curve of liquid bromine, which has been sketched in from the data of Porret (1937).

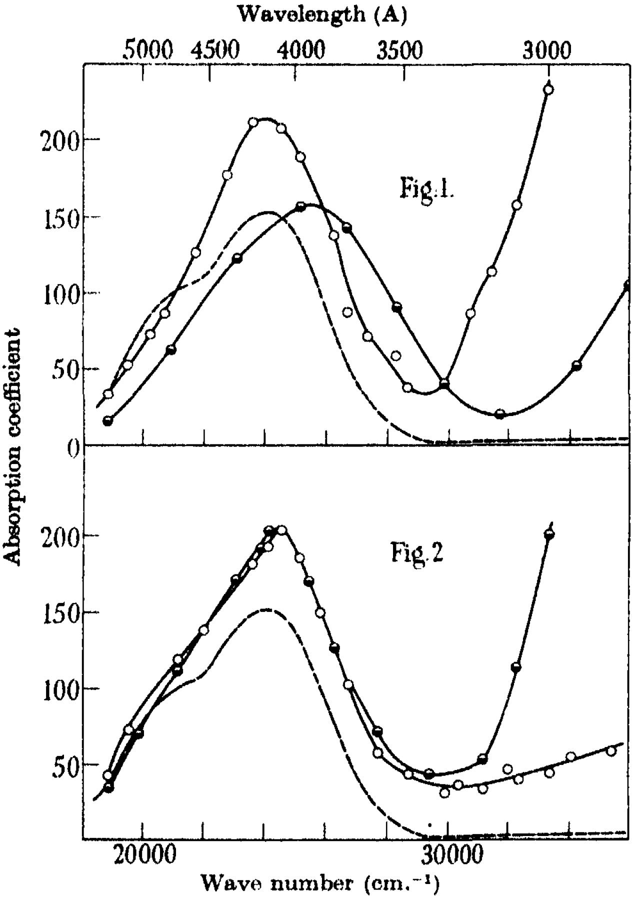


Fig. 1. Broken curve, gaseous bromine; ○ bromine dissolved in cyclohexane; ● bromine dissolved in 2N sulphuric acid.

Fig. 2. Broken curve, gaseous bromine; ○ bromine dissolved in carbon tetrachloride; ● bromine dissolved in chloroform.

Previous work on the absorption spectrum of bromine in solution includes that of Bovis (1924, etc.), who studied the solvents water, carbon tetrachloride, chloroform, ethanol and carbon bisulphide, and of Gillam and Morton (1929), who studied the absorption of halogens dissolved in carbon tetrachloride. Gillam and Morton present their results only in the form of curves; but they appear to be in reasonably good agreement with ours. The results of Bovis are quoted in terms of optical densities, which on recalculation give absorption coefficients that are mostly lower than ours.

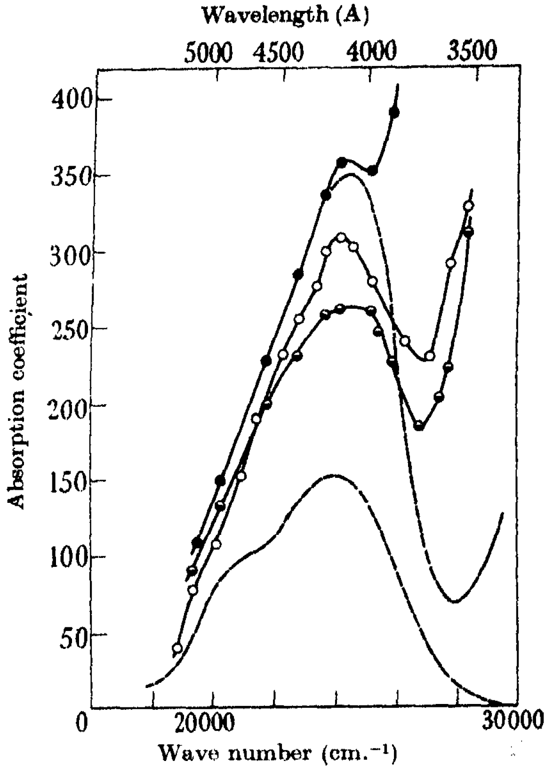


Fig. 3. Upper broken curve, liquid bromine; lower broken curve, gaseous bromine; ○ bromine dissolved in benzene; ◒ in chlorobenzene; ● in toluene.

Discussion

In general, the effect of environment on the absorption of a substance may appear as a bodily displacement of the absorption, or as a change in the absorption coefficient (i.e. of the transition probability), or as a combination of both. In the case of bromine, it is necessary to consider these possibilities with respect to each of the several components of the absorption. The two main components in the absorption of gaseous bromine, A ($\lambda_{\max} = 4150$ A) and B ($\lambda_{max} = 4950$ A), are quite distinct in the broken curves of figs. 1–3. In addition to these, there is the absorption throughout the quartz ultra-violet, ϵ rising from a minimum of 0·2 at 3036 A to about 3 at 2165 A. This absorption is probably complex, and may contain weak bands (Aickin and Bayliss 1938).

The ultra-violet absorption. The most obvious feature of figs. 1–3 is the tremendous increase in the intensity of the ultra-violet absorption of bromine in solution, as compared with the gaseous state. The most marked effect occurs with the aromatic solvents, while liquid bromine occupies a place intermediate between these and the non-aromatic solvents. Another feature common to all the solutions in this region is the occurrence of deviations from Beer's law, ϵ increasing slightly with the concentration. These deviations are considerable in the case of carbon tetrachloride solutions, where the irregularity of the absorption curve above 30,000 cm.$^{-1}$ indicates the possible presence of diffuse band absorption in this region. A similar effect was noticed by Aickin and Bayliss in gaseous bromine at rather higher frequencies. Such deviations from Beer's law set a limit to the accuracy of absorption coefficients in this region, and probably explain the differing results of various authors.

In several solvents, it was possible to find a maximum in the ultra-violet absorption, when by using thin films of solution (about 1 mm.) in the Baly tube we were able to reduce the absorption due to the solvent itself. In each case the maximum due to the dissolved bromine was quite distinct from the solvent cut-off, and the positions of the maxima are given in Table II.

TABLE II. ULTRA-VIOLET MAXIMA IN BROMINE SOLUTIONS

	Absorption maximum in A due to	
Solvent	Bromine	Solvent
Cyclohexane	2500	2150
Chloroform	2725	2230
Benzene	2925	2600
Chlorobenzene	2960	2650
Toluene	2975	2750

Bovis (1927) observed a maximum in chloroform solutions at 2700 A, but we were unable to find the maxima that he records at 2580 A in both water and carbon tetrachloride solutions.

It does not seem possible that these maxima could be due to compounds formed by chemical reaction between bromine and the solvent. In the case of the aromatic solvents, for instance, the absorption of the various bromobenzenes and bromotoluenes is recorded as beginning at about 2800 A and extending to shorter wave-lengths (Purvis 1915; Ley and Engelhardt 1910). There is no maximum in the ultra-violet absorption of liquid bromine as far as 2500 A, or probably even as far as 2200 A (Porret 1937). In gaseous bromine, the maximum may have been reached at 2200 A (Aickin and Bayliss 1938), although the measurements do not extend quite far enough to be certain. However, both Porret and the latter authors agree in considering the ultra-violet absorption of bromine to be complex, and in support of this, unpublished measurements by Bayliss and Rees show a definite maximum at about 2700 A when bromine is mixed with carbon dioxide. This coincides with the maximum that is obtained in chloroform solution. The maxima in other solvents are as yet unexplained, unless they are the same maximum which has been displaced by the solvent effect. Such large displacements seem unlikely in view of the small displacements that are suffered by the *A* and *B* maxima in the same solvents.

The ultra-violet absorption of iodine is known to be greatly affected by solvents, and here too, deviations from Beer's law have been observed (Groh and Papp 1930).

The visible absorption continuum ($\lambda_{\max} = 4150$ A). The effect of solvents on the visible continuous absorption of bromine is to produce a marked increase in the absorption intensity accompanied by a small or zero shift of the maximum towards shorter wave-lengths, except in the case of water, where there is a large displacement of the maximum ($\Delta\nu = 1750$ cm.$^{-1}$), but only a small increase in the intensity.

The total solvent effect is the sum of the effects that are exerted separately on the *A* and *B* components of the continuum. From our results, it is not possible to obtain the effect on the *B* continuum with any accuracy. In gaseous bromine, Acton, Aickin and Bayliss (1936) obtained the absorption curve of the *B* continuum in the following way. They assumed that when plotted against wave numbers as abscissae, the absorption curve of the *A* continuum was symmetrical about its maximum, and that the contribution of the *B* component was practically zero at the maximum of *A*. The absorption coefficients on the high-frequency side of the *A* maximum were then used to determine the contribution of *A* to the absorption on the low-frequency side, and the remainder of the absorption was attributed to the *B* component. This procedure cannot be applied to the absorption in solution, since in the important region of the high-frequency side of the *A* component, there is a considerable but unknown contribution from the ultraviolet absorption, making it impossible to disentangle the *B* component with any accuracy.

It is evident, however, from figs. 1–3 that the displacement of the *B* maximum in the non-aqueous solvents is small, if not zero. The intensity also appears to have been affected less than in the case of the *A* component. In cyclohexane, the intensity of the *B* components is actually less than in the gaseous state.

In the case of the *A* component, the increase in intensity is about 30% in chloroform and carbon tetrachloride, and about 45% in cyclohexane. At first sight, the effect of the aromatic solvents seems to be much greater; but in these cases the absorption at the *A* maximum receives a large contribution from the enormously enhanced ultra-violet absorption. Taking this into account, the net effect in the aromatic solvents is of about the same magnitude as in the other non-aqueous solvents, and in no case is the main absorption enhanced to the same extent as in liquid bromine.

In cyclohexane, the *A* maximum is not displaced relative to the gaseous state. In the other non-aqueous solvents, there are small displacements to higher frequencies, the greatest shift being 550 cm.$^{-1}$ in carbon tetrachloride. Scheibe (1937) has interpreted such shifts in terms of displacements of the potential energy curves of the molecular states that are concerned in the absorption. If his explanation applies in the present case, any such displacements are small.

However, we believe that another explanation must be considered, in view of the possible complexity of the *A* continuum (Mulliken 1936). If the *A* continuum consists of two components of about equal intensity, and which are not exactly superimposed, and if the effect of the solvent is to increase the intensity of the high-frequency component more than the other, then the absorption maximum will appear to shift to higher frequencies. Such an explanation is supported by the absorption curves in figs. 2 and 3, particularly in the solvents carbon tetrachloride, chloroform and benzene. The displacement of the maximum has been accompanied by a definite sharpening, and in our detailed results in Table I there is definite evidence of the appearance of a shoulder just on the low-frequency side of the maximum.

The interpretation of the various phenomena suffers from the uncertainty that still exists in the theory of the bromine continua. Mulliken (1936) has discussed the following alternative assignments of various electronic transitions to the *A* and *B* components of the absorption:

(1) *A*: mixture of ${}^1\Pi_u \leftarrow {}^1\Sigma_g^+$ and ${}^3\Pi_{0+u} \leftarrow {}^1\Sigma_g^+$.
B: ${}^3\Pi_{1u} \leftarrow {}^1\Sigma_g^+$.
Coupling: $\Omega - s$.

(2) A: ${}^1\Pi_u \leftarrow {}^1\Sigma_g^+$.
B: mixture of ${}^3\Pi_{0+u} \leftarrow {}^1\Sigma_g^+$ and ${}^3\Pi_{1u} \leftarrow {}^1\Sigma_g^+$.
Coupling: approaching $\Lambda - \Sigma$.

He regards (1) as more probable on general grounds, the main evidence against it being that of the Franck-Condon principle. The transition to ${}^1\Pi_u$ is thought to lie at slightly higher frequencies than that to ${}^3\Pi_{0+u}$. Explanation (1) is in accord with our interpretation of the displacement of the A maximum in non-aqueous solvents. The effect of solvents on the intensities of A and B is interpreted as showing that the transition probabilities are affected in the decreasing order ${}^1\Pi_u \leftarrow {}^1\Sigma_g^+$, ${}^3\Pi_{0+u} \leftarrow {}^1\Sigma_g^+$, ${}^3\Pi_{1u} \leftarrow {}^1\Sigma_g^+$.

(2) receives its greatest support from the Franck-Condon principle (Darbyshire 1937; Bayliss 1937). The A maximum is considered to be single, and the displacement of the maximum must be due to displacement of potential energy curves as discussed by Scheibe. Our results would then indicate that solvents have a greater influence on the ${}^1\Pi_u \leftarrow {}^1\Sigma_g^+$ transition than on the ${}^3\Pi_u \leftarrow {}^1\Sigma_g^+$ transitions. But (2) corresponds to assuming a coupling that approaches $\Lambda - \Sigma$, where inter-system transitions are forbidden. Under these circumstances, one would expect the "forbidden" transitions ${}^3\Pi_u \leftarrow {}^1\Sigma_g^+$ to be more affected by the perturbing influence of solvents than ${}^1\Pi_u \leftarrow {}^1\Sigma_g^+$, which is the reverse of the experimental observation. The phenomena in solution thus seem to favour (1) as the correct interpretation of the continua. Although the Franck-Condon evidence in favour of (2) is strong, it is however based on the assumption that the Morse curve of the ${}^3\Pi_{0+u}$ state of bromine is correct up to high vibrational quantum numbers.

The large displacement of the absorption in aqueous solution must be interpreted according to Scheibe (1937). It will be noticed (fig. 1) that the absorption as a whole has been broadened, and that the individuality of the B component has been lost. The A component has been increased little in intensity, and it is difficult to say whether the intensity of B has been decreased, or whether it has been merged in A by virtue of a greater displacement towards higher frequencies. Aqueous solutions of bromine recall the phenomena that have been observed in aqueous and alcoholic solutions of iodine, where a large displacement of the maximum towards the ultraviolet is accompanied by an actual decrease in the intensity (Getman 1928; Groh and Papp 1930).

SUMMARY

The visible and ultra-violet absorption spectrum of bromine has been investigated in the following solvents: water (2N sulphuric acid), cyclohexane, chloroform, carbon tetrachloride, benzene, chlorobenzene, toluene.

The weak ultra-violet absorption of gaseous bromine is enormously enhanced in solution, particularly by the aromatic solvents. The smallest effect occurs in carbon tetrachloride, where considerable deviations from Beer's law were found. Absorption maxima in the ultra-violet were found at 2500 A in cyclohexane, at 2725 A in chloroform, and in the neighbourhood of 2950 A in the aromatic solvents.

In the region of the visible continuum, the non-aqueous solvents cause a marked increase in absorption intensity combined with a zero or small shift of the maximum towards the ultra-violet. Water causes a large displacement (1750 cm.$^{-1}$) of the maximum to higher frequencies; but only a very small increase in intensity. If the *A* component of the visible continuum is complex, the displacement of the maximum in non-aqueous solvents can be interpreted as due to a greaters olvent effect on ${}^1\Pi_u \leftarrow {}^1\Sigma_g^+$ than on ${}^3\Pi_{0^+u} \leftarrow {}^1\Sigma_g^+$. The transition ${}^3\Pi_{1u} \leftarrow {}^1\Sigma_g^+$ is less affected still. Difficulty is found in interpreting the observations if the *B* continuum is complex and *A* simple.

REFERENCES

Acton, A. P., Aickin, R. G. and Bayliss, N. S. 1936 *J. Chem. Phys.* **4**, 474.
Aickin, R. G. and Bayliss, N. S. 1938 In the Press.
Bayliss, N. S. 1937 *Proc. Roy. Soc.* A, **158**, 551.
Bovis, P. 1924 *C.R. Acad. Sci., Paris*, **178**, 1964.
— 1927 *C.R. Acad. Sci., Paris*, **185**, 57.
— 1928 *Ann. Phys., Paris*, **10**, 232.
Darbyshire, O. 1937 *Proc. Roy. Soc.* A, **159**, 93.
Getman, F. H. 1928 *J. Amer. Chem. Soc.* **50**, 2883.
Gillam, A. E. and Morton, R. A. 1929 *Proc. Roy. Soc.* A, **124**, 604.
Groh, J. and Papp, S. 1930 *Z. phys. Chem.* A, **149**, 153.
Harrison, G. R. 1929 *J. Opt. Soc. Amer.* **19**, 267.
Ley, H. and Engelhardt, K. 1910 *Z. phys. Chem.* **74**, 1.
Liebhafsky, H. 1934 *J. Amer. Chem. Soc.* **56**, 1500.
Mulliken, R. S. 1936 *J. Chem. Phys.* **4**, 620.
Porret, D. 1937 *Proc. Roy. Soc.* A, **162**, 414.
Purvis, J. E. 1911 *J. Chem. Soc.* **99**, 1699.
— 1915 *J. Chem. Soc.* **107**, 496.
Rabinowitch, E. and Wood, W. C. 1936 *Trans. Faraday Soc.* **32**, 540.
Scheibe, G. 1937 *Angew. Chem.* **50**, 212.
Twyman, F. and Allsopp, C. B. 1934 "Spectrophotometry." Hilger.
Weissberger, A. and Proskauer, E. 1935 "Physical Constants of Organic Solvents." Oxford.

The development of the spark discharge. II

By T. E. Allibone, D.Sc., Ph.D. and J. M. Meek, M.Eng.

(*Communicated by S. R. Milner, F.R.S.—Received* 26 *August* 1938)

[Plates 10–16]

1. Introduction

An account has already been given of observations made with a rotating-film camera in the study of the electric spark in air at atmospheric pressure (Allibone and Meek 1938; Allibone 1938). The object of the present paper is to extend the previous results, and in particular to describe the development of the spark in air at reduced pressures, viz. from 76 to 0·3 cm. Hg. The high-speed cathode-ray oscillograph has been used in conjunction with the rotating camera to enable voltage and current oscillograms to be made. Various phenomena, previously observed by oscillographic methods only, have now been shown to be associated with definite processes in the discharge mechanism.

2. Apparatus and technique

The general features of the apparatus used in the present work are no different from those previously described (Allibone and Meek 1938;* referred to later as (I)). The same 2000 kV impulse generator has been used throughout the experiments, though for the low-pressure work only two of the twelve stages of condensers were used; these had a series capacitance of 0·06μF and gave a maximum output voltage of 330 kV. One terminal of the generator was connected to the discharge gap through series resistances of different values; the other terminal, together with one electrode of the discharge gap, were earthed.

The high-voltage electrode used in the experiments at atmospheric pressure consisted of a brass rod 1 cm. in diameter with a point tapered to a cone of 30°. This electrode was suspended vertically over a plane earthed electrode at heights varying from 25 to 200 cm. The plane electrode (240 × 240 cm.) was of sufficient extent to prevent discharges terminating at its edge.

---

* Erratum in Paper I: page 123, Table I, column 2. Time-lags for gap lengths of *a* above 25 cm. should be $\times 10^{-6}$ sec.

In the experiments at reduced pressures the discharge gap was mounted in a glass cylinder 37 cm. internal diameter and 100 cm. in length. One end was sealed by a bakelite board, 1 in. thick, the other end by a steel plate; the chamber was evacuated by a rotary oil pump and the pressure was recorded on a mercury manometer. The high-voltage electrode consisted of a brass rod, $\frac{3}{4}$ in. in diameter, with a steel tip in the form of a 30° cone; this electrode, suspended from the bakelite board, projected 28 cm. into the cylinder coaxially, so that the development of the discharge from the electrode tip is unlikely to have been influenced by any discharge over the surface of the bakelite. With regard to the earthed electrode, it had been the authors' intention to use a plane in order to obtain complete dissimilarity between the electrodes. It was found, however, that for the selected gap of 50 cm. the negative discharge frequently terminated on the edge of the plane and also took place over the inner surface of the glass cylinder in the vicinity of the plane even though the edge of the plane was turned to a radius of 5 cm. A short pointed steel rod 1 cm. long placed in the centre of the earthed electrode had the effect of concentrating the negative discharges on to its tip and almost eliminated sparkover on the surface of the glass, whilst it had no appreciable effect on the behaviour of the positive discharge. One half of the inside surface of the glass cylinder was covered with black paper in order to prevent confusion of the photographs of the discharge due to reflexion from this surface. The air in the chamber was frequently changed, so that the character of the discharge would not be affected by change in gas composition.

The rotating-film camera described in the previous paper (I) was again used in the experiments; the only alteration was the introduction of a new quartz-rocksalt lens ($f/4{\cdot}5$: focal length 12 cm.) which considerably enhanced the quality of the photographs. The special value of the transmission of the far ultra-violet radiations by this lens was of course lost in the low-pressure experiments as a glass tank had to be used.

Records of the voltage applied to the discharge gap, and also of the current in the discharge, were obtained by means of a high-speed cathode-ray oscillograph. In some cases a capacitance potential divider was connected in parallel with the discharge gap and coupled to the oscillograph. In other cases the voltage applied to the oscillograph was derived from an antenna arranged close to the high-voltage electrode. This antenna was carefully shielded from the earthed electrode to reduce the influence of upward-developing streamers on the recorded voltage. The latter method is admittedly less accurate than the former method so far as the voltage ordinate is concerned, but the use of a condenser potential divider considerably modifies the wave-front of the applied voltage when the series resistance is large and is therefore undesirable. The current in the discharge was measured by connecting the lower electrode to earth through a non-inductive resistance, the value of which was sufficient to enable the voltage developed across it to be applied directly to the oscillograph.

3. GENERAL DESCRIPTION OF RECORDS

The nomenclature described in the previous paper (I) has again been adopted to describe the different processes in the development of the discharge. The stepped leader process already observed has been found to be accentuated when the discharge takes place in air at less than atmospheric pressure, with a resultant decrease in the speed at which the envelope of the tips of the individual leaders progresses. Neither the speed of the main stroke, nor that of the leader stroke traversing a pre-ionized path, has yet been measured but is known to be in excess of 3×10^8 cm./sec. (I). Thus in the analysis of the photographs the main stroke is again used as the basis for the measurement of leader stroke velocity.

In all the illustrations of the discharges the high-voltage electrode is at the top, and the earthed electrode at the bottom, the leader stroke appears on the left, the main stroke on the right. The oscillograms given in the text are printed with their time axes from left to right to correspond to the photographs of the sparks: a zero line is given for reference, and the natural period of the timing wave is given on each record. The spacing d in cm. between the electrodes, the circuit resistance R in ohms, the pressure P in cm. Hg. and, where relevant, the height p of the earthed point above the earthed plane are given on each photograph. In all the records the minimum voltage required to cause sparkover was used (unless otherwise stated).

As there is such a very marked difference between the discharges when the polarity of the high-voltage electrode is changed, detailed description of the results will be separated into two categories, positive and negative discharges, each being subdivided into sections, atmospheric and sub-atmospheric pressures.

4. THE POSITIVE DISCHARGE

(a) *Atmospheric pressure*

Considerable attention has already been given to the positive discharge (I) and relatively little can be added to the previous details. However, as the technique of voltage measurement involved the use of a condenser voltage divider placed across the gap, the effect of the presence of this divider on the development of the discharge has been investigated. It will be realized that the capacitance C in parallel with the gap cannot be altered without at the same time altering the rate of rise of applied voltage unless the series resistance is also altered by a proportionate amount. Experiments were therefore made with varying values of C with the same value of R, and then with different values of R keeping CR approximately constant. The results can be expressed in the following manner:

(A) Addition of capacitance increases the thickness of the photographic record of the main stroke, i.e. the current in the stroke is greater.

(B) Addition of capacitance in so far as it is accompanied by a greater value of CR increases the number of stepped leader strokes—these steps begin to occur on the rising front of the voltage wave.

(C) Addition of resistance in so far as it is accompanied by a greater value of CR increases the number of stepped leader strokes as in (*b*) above: note when C is very small and R is large the current in the leader stroke causes a drop in the terminal voltage and the voltage waveform takes on a "saw-tooth" appearance.

(D) Addition of capacitance, keeping CR constant, increases the current flowing in the final stepped leader stroke, and the current is maintained at that higher value until the main stroke occurs. In the limit as R reaches low values the whole picture is too blurred by halation from the main stroke to enable deductions to be made.

Records illustrating the above are given in fig. 1, Plate 10, for a gap of 100 cm. (*a*) with a resistance of 100,000 ohms, and capacitances of 10, 200 and 5000 $\mu\mu$F, and (*b*) with a capacitance of 10 $\mu\mu$F, and resistances of 10,000, 100,000 and 1,000,000 ohms. In fig. 1 (*a*) series, the wave-front of the applied voltage was approximately 2 μsec. for the first record (i), and it will be observed that a single leader stroke traverses the whole gap in 46 μsec., and is then followed by the main stroke: no partial discharges precede this leader stroke, which has a final velocity of $2 \cdot 5 \times 10^6$ cm./sec. In the second record (ii) the wave-front was approximately 50 μsec., and during this time several stepped leader strokes develop before the final leader stroke traverses the gap. This leader stroke, which is initiated 13 μsec. before the main stroke occurs, has a final velocity of $3 \cdot 6 \times 10^6$ cm./sec. In the third record (iii) there are stepped leader strokes (not visible in the printed record) for 140 μsec. before the final leader stroke develops (this is blurred by the main stroke). The main stroke is very heavy and is followed by multiple strokes which occur after a further 100 and 210 μsec.

respectively. These multiple strokes are similar to those examined by Malan and Collens (1937) in the case of the lightning flash.

In the fig. 1 (*b*) series, the first record (i) shows a weak initial leader stroke followed by a strong leader the intensity of which is maintained until the main stroke occurs after 12 μsec. This latter leader stroke is strongly branched and the subsequent retracing of these branches by the main stroke is most evident. The "final" velocity of the leader after it has traversed 75% of the gap is $8 \cdot 10^6$ cm./sec. The second record (ii) is of course comparable with (i) of the first series: the leader traverses the whole

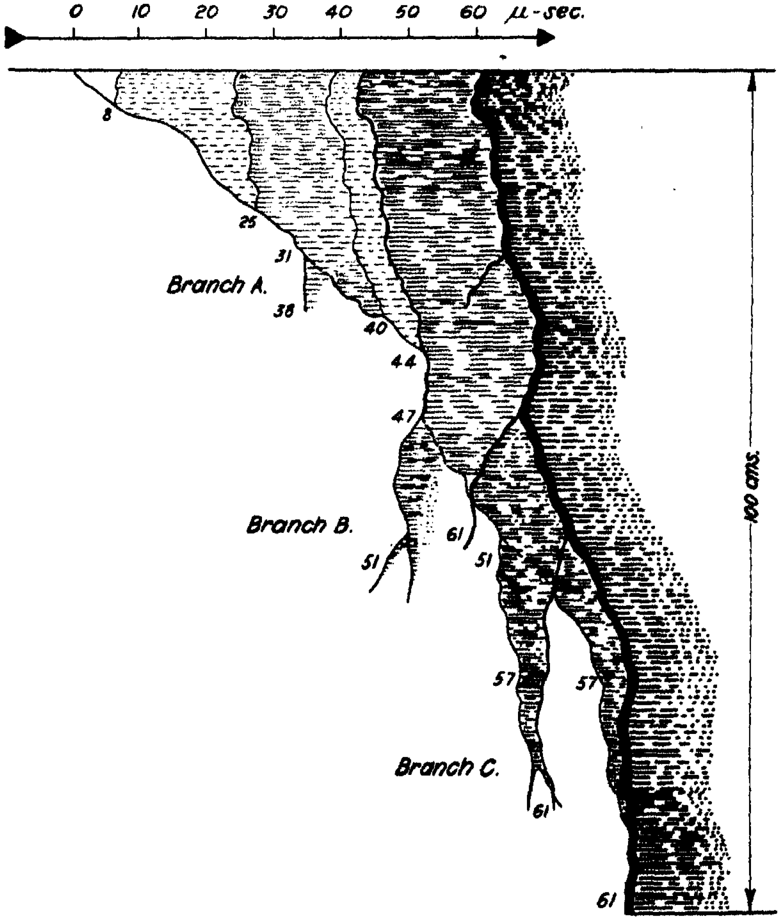


FIG. 2. Positive point electode: earthed plane electrode. Drawing based on an original photograph of a discharge over a 100 cm. gap. $R = 100{,}000$; $C = 300$. (A photograph of a similar discharge is given in fig. 1 (*a*) (ii).)

gap in 46 μsec. with a final velocity of $2\cdot5 \times 10^6$ cm./sec. In the third record (iii) four well-defined stepped leaders occur at intervals of 14, 22 and 41 μsec., the last develops nearly to the opposite electrode with a final velocity of $2\cdot4 \times 10^6$ cm./sec. and meets a short negative leader 7 cm. above the earthed plane. The occurrence of this type of discharge was mentioned in (I) (p. 101), but no record was then available from which the speed of the negative leader stroke could be measured: in this record the speed is 2×10^6 cm./sec. This upward developing leader is also branched.

It will be seen from the above that the detailed formation of the leader stroke can be altered by changing the circuit constants; their effect on the speed of formation of the leader stroke has already been studied in (I).

As mentioned above, the addition of capacitance increases the current and luminosity of the stepped leader stroke and details of the branching of the leader stroke are thereby enhanced. A number of records of discharges under these conditions exhibit the type of branching described by Schonland, Malan and Collens (1935, p. 598). These are illustrated in fig. 2, a drawing of a discharge 100 cm. long, $R = 100{,}000$, $C = 200$ μμF. The numbers to the left of the leader stroke and its branches denote the times in microseconds subsequent to the initiation of the discharge. The branches A and B have ceased to develop, whereas the branch C is still developing when the main stroke occurs and follows the paths of each of the branches exactly as described for the lightning flash.

(*b*) *Subatmospheric pressures*

Records of spark propagation have been taken with three different values of series resistance and at pressures from 76 cm. to 0·3 cm. Hg. The description of records will be made in order of increasing resistance. Fig. 3, Plate 11, gives representative records: *a*, $R = 700$ ohms; *b*, $R = 10{,}000$ ohms; *c*, $R = 100{,}000$ ohms: and in each group, (i) is at 40 cm. pressure, (ii) at 20 cm. pressure, (iii) at 10 cm. pressure, and (iv) at 5 cm. pressure.

With low resistance, halation due to the main stroke obliterates the leader stroke at atmospheric pressure: at 40 cm. pressure (fig. 3 (*a*) (i)) the leader stroke can just be distinguished; at 20 cm. pressure (fig. 3 (*a*) (ii)) stepped leader strokes become very pronounced on the principal channel and on the branches; at 10 cm. pressure (fig. 3 (*a*) (iii)) an ill-defined brush discharge from the anode is followed after very long time intervals (50–70 μsec.) by isolated stepped leader strokes and finally, after about 100 μsec. the main discharge occurs; at 5 cm. pressure (fig. 3 (*a*) (iv)) the brush discharge from the anode becomes even more diffuse, and extends further towards the opposite electrode; its speed of development cannot be measured on account of its lack of definition. It is accompanied by a small discharge at the cathode which may occur almost at the same instant as the start of the anode brush. A slow speed leader stroke then develops from the anode along the core of the diffuse brush discharge; the speed is estimated to be 2×10^5 cm./sec. near the anode but increases as it meets the ascending brush discharge. The main stroke then follows. With series resistance of 10,000 ohms (fig. 3 (*b*)) the leader stroke and stepped leader strokes are well developed at 76, 40, 20, 10 and 5 cm. pressure. Record fig. 3 (*b*) (ii) very clearly shows the development of nine stepped leader strokes, each branched at their tips. The discharge paths at 5 cm. Hg (fig. 3 (*b*) (iv)) are not so diffuse as in fig. 3 (*a*) (iv), though each step becomes ill-defined at its tip. With series resistance of 100,000 ohms (fig. 3 (*c*)) the stepped leader strokes are very clearly defined at all pressures.

At pressures lower than 5 cm. Hg records could only be taken for $R = 700$ ohms, fig. 4 (*a*), Plate 12: and for $R = 10{,}000$, fig. 4 (*b*), Plate 12. For $R = 100{,}000$ ohms the discharge paths were too weak to be identified. Rotating camera records were taken at 2·5 and 1·0 cm. Hg and stationary camera records at 0·9, 0·6 and 0·35 cm. Hg. The discharge paths are always diffuse at these pressures and the leader stroke travels apparently in a continuous and not a step-by-step method with a speed of about 2×10^5 cm./sec.: it is not branched. Rotating camera records of discharges at less than 1 cm. Hg were unintelligible on account of the diffuseness of the discharge. Three stationary camera records are given in fig. 5, for $R = 700$ ohms. At 0·35 cm. Hg a diffuse pink glow discharge spreads out from the anode to fill the chamber as shown in fig. 5 (*a*), Plate 12. A cathode dark space is evident, and also patches of light on the cathode due to positive-ion bombardment. With increase in pressure to 0·6 cm. Hg a certain concentration of the discharge into a central core is apparent as shown in fig. 5 (*b*), Plate 12, and this core is terminated at the anode by an intense streamer 5–10 cm. in length, whilst at the cathode the core appears to concentrate on several small tufts of discharge. The discharge at 0·9 cm. Hg as illustrated in fig. 5 (*c*), Plate 12, shows a streamer developing from the cathode in conjunction with the streamer from the anode; the whole chamber is filled with a pink glow, and a more concentrated column of glow discharge is seen to join the two streamers. For discharges at pressures lower than 1 cm. Hg complete sparkover in the accepted sense did not occur with this polarity: an increase of applied voltage does not cause an increase in the length of the streamers beyond a certain value but only an increase in their intensity and in the intensity of the column of glow discharge joining their tips.

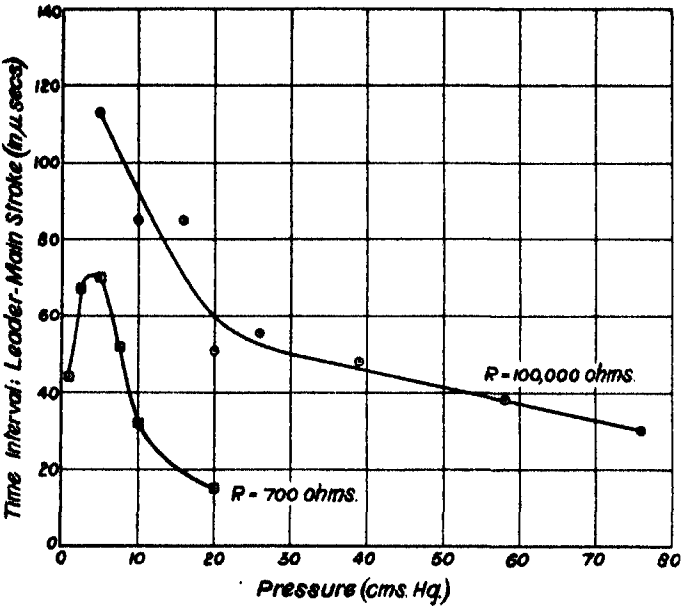


FIG. 6. Positive point electrode: earthed plane ($p = 1$ cm.) electrode. Curves relating the time interval between the initial positive leader and main stroke with pressure for a 50 cm. discharge gap.

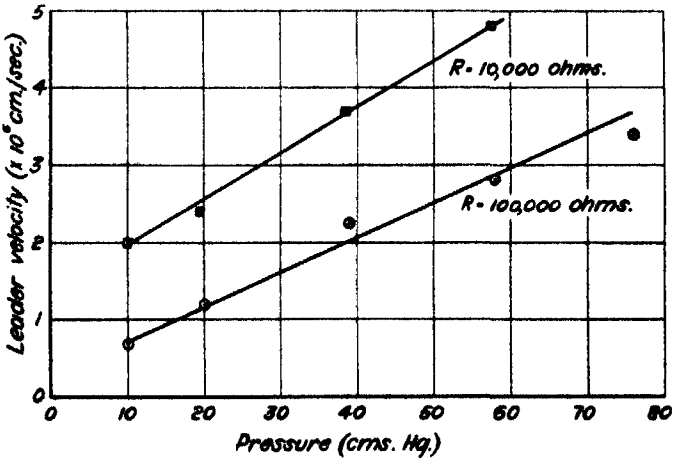


FIG. 7. Positive point electrode: earthed plane ($p = 1$ cm.) electrode. Curves relating leader stroke velocity at 70 % gap length with pressure for a 50 cm. discharge gap.

It has already been noted in (I) that the time interval between leader and main stroke increases with increasing value of circuit resistance: this feature persists down to very low pressures, and in addition the time intervals increase with diminishing pressure for any one value of series resistance. Curves for these time intervals are given in fig. 6 for two values of series resistance, 700 and 100,000 ohms. As the pressure is diminished the deviation of results from the mean values increases and may reach

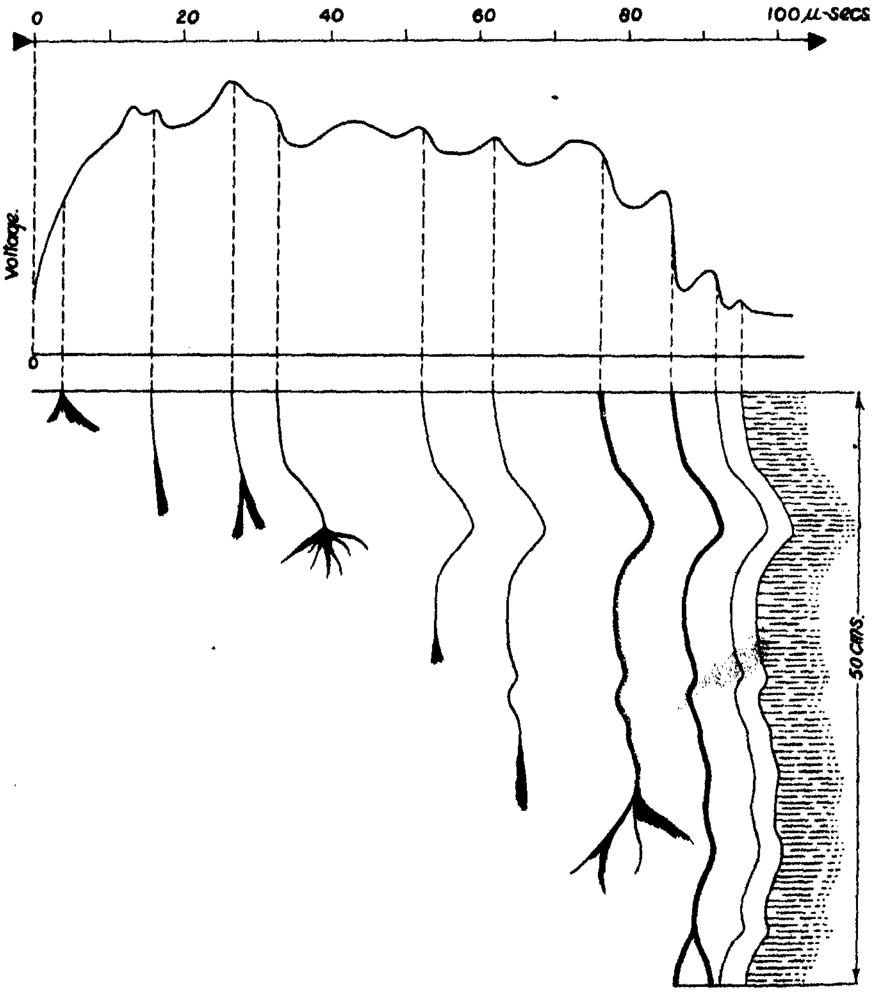


FIG. 9. Positive point electrode: earthed plane ($p = 1$ cm.) electrode. Drawing based on an original photograph of a discharge and the synchronized voltage oscillogram. $R = 100{,}000$: $P = 10$.

50 % at pressures of a few cm. Hg. Very many photographic and oscillographic records were taken therefore to obtain representative values of the leader/main stroke time interval and there can be no doubt as to the reality of the maximum value of this interval as recorded in the lower curve: no explanation can be advanced for its occurrence. The variation of the velocity of the leader stroke with series resistance and gap length has also been discussed in (I). A further variation with pressure has been observed

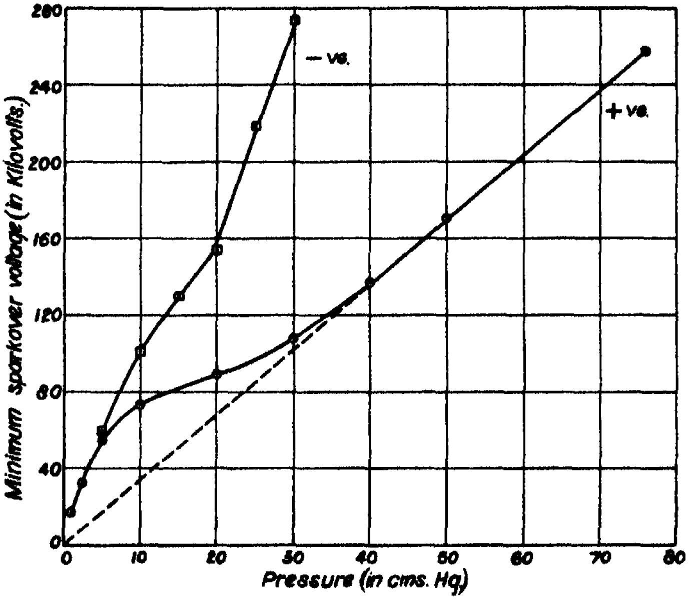


FIG. 10. Relation between minimum sparkover voltage and pressure for a 50 cm. discharge gap between a high-voltage point and an earthed plane ($p = 1$ cm.).

and curves are given in fig. 7 which show a more or less linear increase of leader stroke velocity with increasing pressure.

The method of recording the voltage at the upper electrode by an antenna and an oscillograph is of course comparable with that used by Appleton and Chapman (1937) and by Schonland (1937). The sensitivity of the method has been increased since it was first used, (I), and with high series resistance the variations of voltage at the terminal are seen to be quite large: they may amount to 20 % for strong leader strokes. An example of such an oscillogram is shown in fig. 8, Plate 13. The time intervals between successive stepped leader strokes are found to correspond with the times between voltage variations picked up on the antenna; a comparison between one of these oscillograms and the synchronized photograph of the discharge is shown in fig. 9 for a series resistance of 100,000 ohms and a pressure of 10 cm. Under these conditions of resistance and pressure no continuous streamer can be observed to join the tips of the individual leader strokes. Calculations based on the observed voltage drop indicate that the current in the leader stroke is of the order of $\frac{1}{4}$ amp.; the value increases with increasing pressure, and this is also confirmed by the increased intensity of the leader stroke at higher pressures. With series resistance of 10,000 and 1000 ohms the voltage variations are proportionately smaller.

The voltage at which complete breakdown occurred was measured by a sphere gap connected in parallel with the discharge gap. The condenser potential divider was in circuit in these tests. It was found that the breakdown/pressure curve—given in fig. 10—did not agree with Paschen's law at pressures below 40 cm. Hg. This divergence may be explained by the long time-lags of breakdown of the gap at low pressures, the lags (see fig. 6) being comparable with the time-to-half-value of the wavetail of the applied impulse (400 μsec.); the impulse voltages necessary to cause breakdown are therefore higher than the A.C. or D.C. voltages with which Paschen's law has hitherto been investigated, the "impulse ratio" therefore increases with diminishing pressure.

5. THE NEGATIVE DISCHARGE

(a) *Atmospheric pressure*

In the previous paper (I) the discharge from a negative high-voltage point to an earthed plane has been studied for gaps no larger than 70 cm. It was however decided to study the point/plane discharge over as large an electrode spacing as possible to see whether the height of the junction point of the negative and positive leaders varied with gap length and series resistance. The resistance was varied down to a value at which the leader stroke was rendered invisible by halation of the main stroke.

It was found that the height of the junction point above the earthed plane expressed as a percentage of the electrode spacing decreased with increase of gap length for otherwise identical conditions. The height of the junction point varied from 56 % of the gap length for a 25 cm. gap to 32 % of the gap length for a 150 cm. gap for $R = 100{,}000$. The height of the junction point was also found to decrease with a decrease in series resistance. For a 50 cm. gap it varied from 53 % of the gap length for $R = 250{,}000$ to 44 % of the gap length for $R = 13{,}000$. The fact that the height of the junction point of the positive and negative leader strokes expressed as a percentage of the gap length diminishes with increasing gap length and decreasing circuit resistance offers a possible explanation of the observed fact that photographs of the negative lightning flash to ground seldom show an upward directed positive leader stroke rising to meet the negative leader stroke.

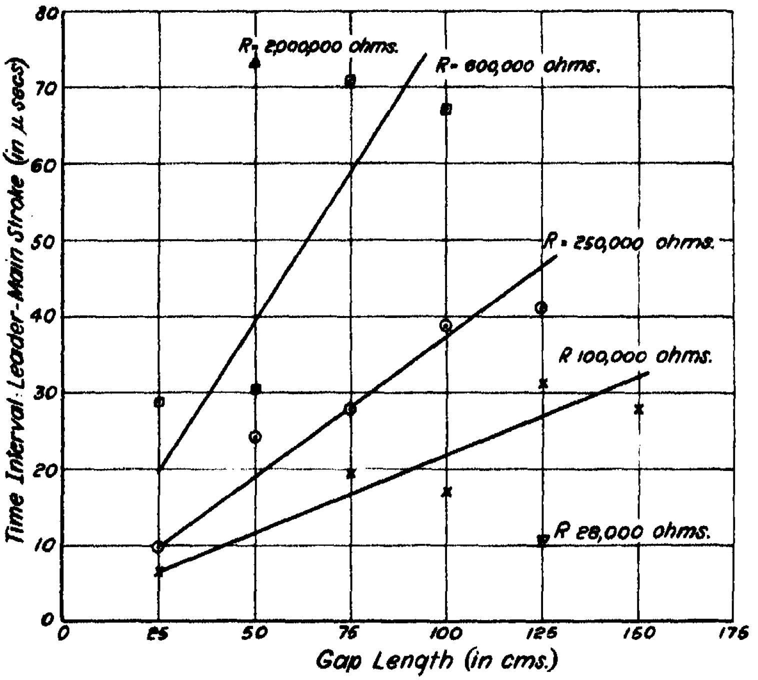


Fig. 11. Negative point electrode: earthed plane electrode. Curves relating the time interval between the initial negative leader and main stroke with gap length for different series resistances.

Further study of the negative discharge over the point/plane gap has confirmed the work previously reported and now curves for different resistances can be given showing the variation with gap length of the time interval between the initiation of the first negative leader stroke and the occurrence of the main stroke. Average values of these time-intervals derived from a number of photographs are given in fig. 11. The curves show an increase of the time interval with gap length and with resistance. The fact that the points do not lie on a uniform curve is due to the wide variation in time interval which may occur, even under identical conditions. The variation may be as great as 50 %, but it is offset to some extent by plotting average values only. The increase of the time interval between the first negative discharge and the main stroke with increasing series resistance can be ascribed—as in the foregoing section on the positive discharge—to the increased time to maximum value of the voltage wave, and to the variation of terminal voltage due to current in the discharge when the resistance is very high.

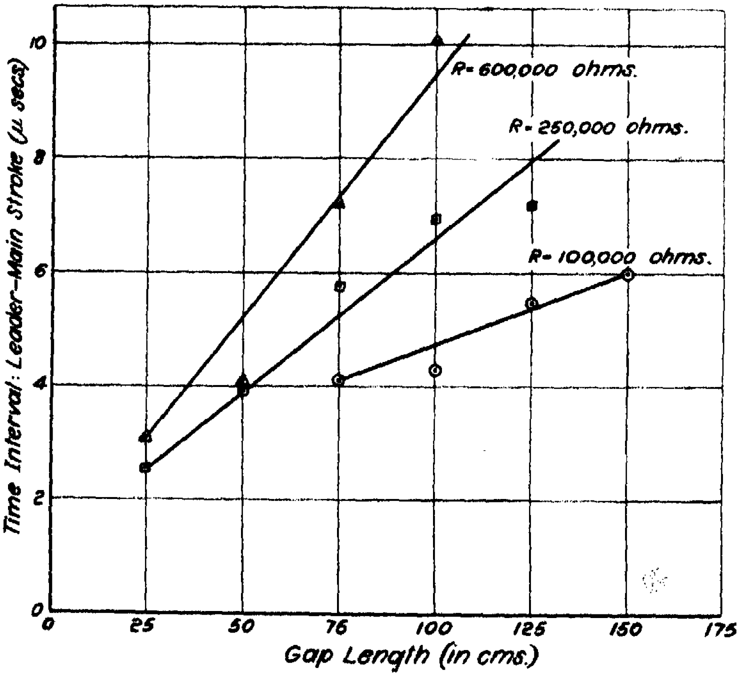


FIG. 12. Negative point electrode: earthed plane electrode. Curves relating the time interval between the initiation of the positive upward-developing leader and the main stroke with gap length for different series resistances.

As previously stated the upward-growing positive leader is generally initiated at the same instant as the final negative leader. Variation of the time interval between the positive leader and the main stroke with gap length and series resistance is shown in fig. 12.

A number of interesting new features of the negative discharge have been observed. In the vicinity of the cathode the discharge often takes the form of thin concentrated filaments some inches in length branched in either direction. Such streamers may be clearly seen on the left of the main spark shown in fig. 13 (*a*), Plate 13, where they sometimes intermingle with branches from the upward-developing positive leader stroke. They may be likened to the mid-gap streamers described by Dunnington

(1931) and to the concentrated filaments which frequently appear in the negative Lichtenberg figures. In fig. 13 (*b*), Plate 13, a diffuse discharge at the cathode becomes concentrated into one of these filaments a few inches away from the cathode and 7 μsec. later a second negative leader stroke follows over the track of the concentrated filament but its connexion to the cathode is over an irregular track.

A photograph of a discharge in which corona and not a complete spark took place is shown in fig. 13 (*c*), Plate 13. The stepped leader process is present at the cathode, while two mid-gap streamers can be clearly seen with their axes collinear and branched in the direction of the cathode. A long branch also extends from the lower mid-gap streamer towards the anode for about 80 % of the gap length.

Another interesting though infrequent phenomenon associated with the negative leader stroke is the appearance of an upward-growing positive type leader in the downward-growing negative leader mechanism. A discharge in which such a leader stroke formation is present is shown in fig. 14 (drawn from a photograph) where a positive leader branching upwards can be seen although the general branching of the discharge in this vicinity is downwards. The path traced out by the positive leader is followed in detail by a subsequent negative leader. Although the speed of the positive leader is too fast to be measured by the present camera for the first few cm. of its path, yet over the last 5 cm. it travels at a speed of 3×10^6 cm./sec. No explanation can yet be provided for the appearance of this positive type leader in the vicinity where it is found.

Variation of series resistance causes an interesting change in the mechanism of the discharge for short point-plane gaps of the order of 25 cm. For resistances up to about 500,000 ohms there is no change in the normal leader mechanism, i.e. negative and positive leaders meet in mid-gap. However, with higher values of resistance the normal positive leader does not develop directly from the plane. A small brush discharge from the cathode is accompanied by a long upward-growing brush discharge from the earthed plane, while from out of this latter discharge, at about 8 cm. above the plane, a leader stroke of the normal positive type develops to meet a negative leader stroke near the cathode, after which the main stroke occurs. Such a discharge is shown in fig. 15 (*a*), Plate 14, for a circuit resistance of 600,000 ohms and the time-interval between the brush discharge and the main stroke is 2·5 μsec. Further increase in the series resistance, to 2 megohms, causes an increase in the height at which the normal positive leader develops from the brush discharge, and in some cases no such positive leader is evident, but the main stroke takes place in a line parallel to the axes of the original brush discharges, as seen in fig. 15 (*b*), Plate 14. This long brush discharge is a feature of the positive discharge from plane-type electrodes: it was referred to in the previous paper (I) for sphere/sphere electrodes but it was not then appreciated that the extent of the brush discharge was a function of the circuit resistance. Recently this same type of positive discharge was seen from the spherical terminal of a very low powered van de Graaff-type electrostatic generator.

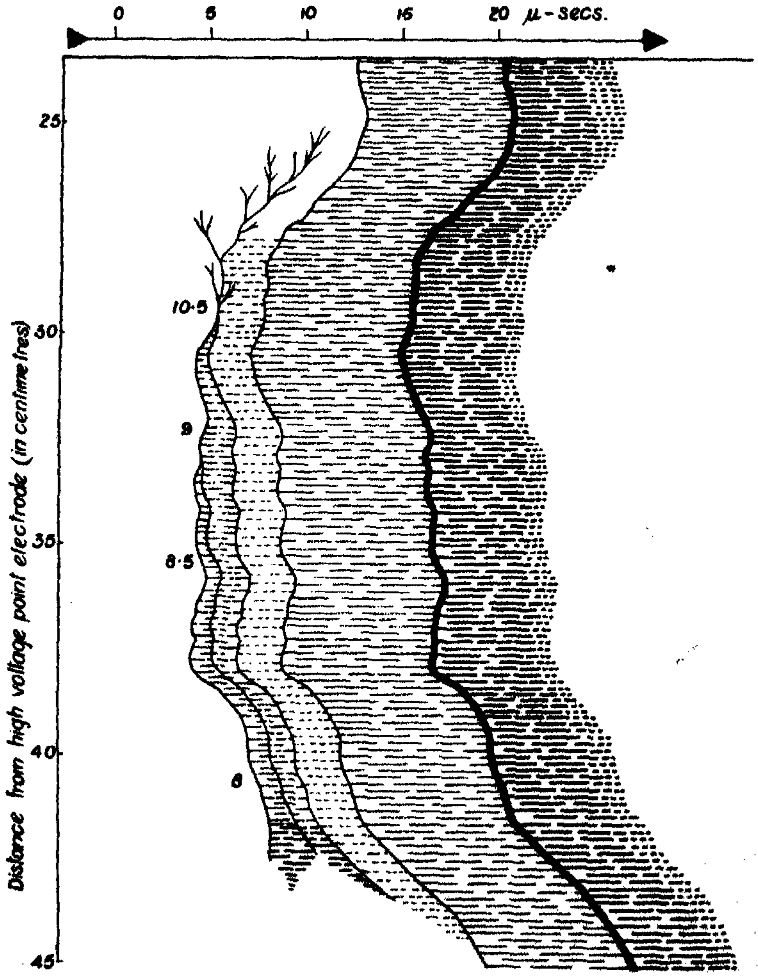
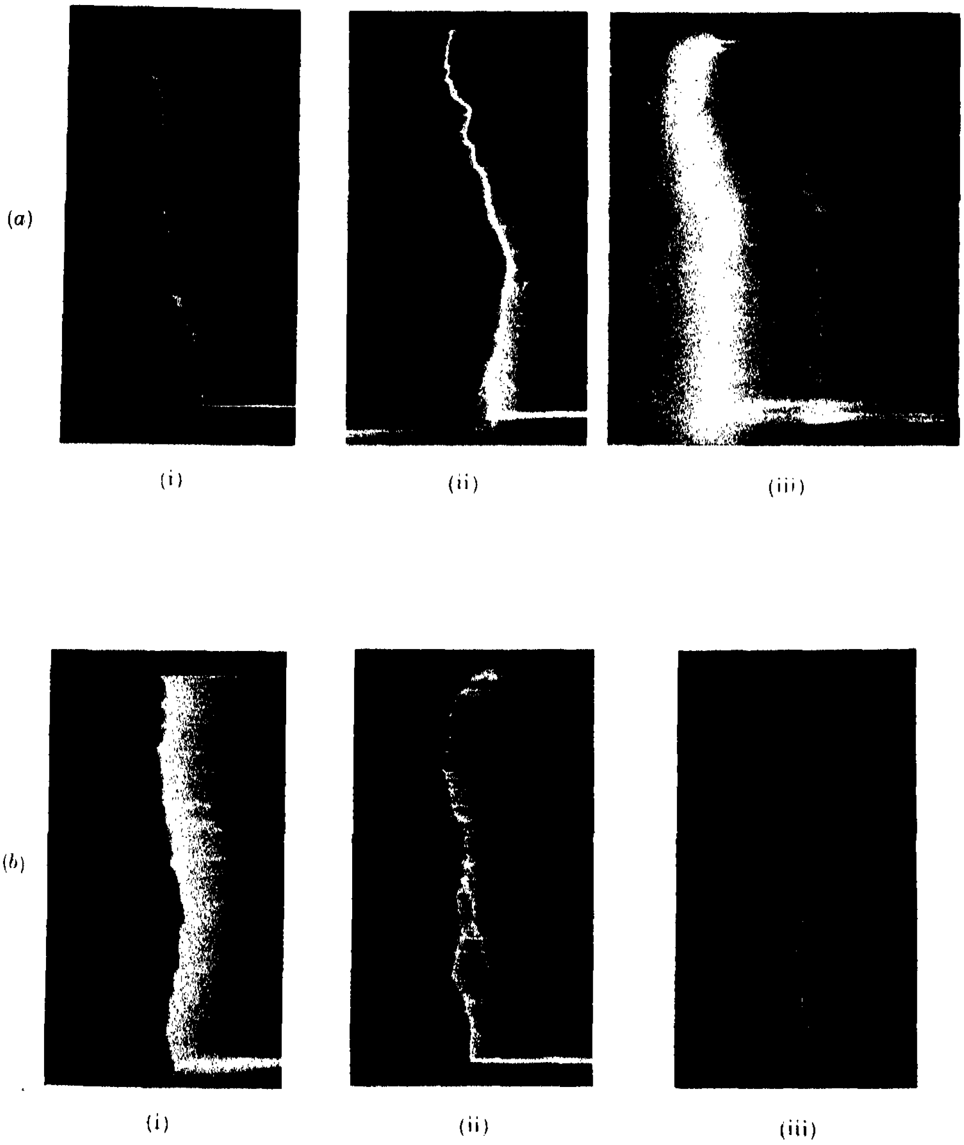


FIG. 14. Negative point electrode: earthed plane electrode. Drawing based on an original photograph to illustrate an upward-developing positive leader in the downward-developing negative leader-stroke mechanism. (The junction point of the negative leader and the main positive leader from the earthed plane is at 70 cm. from the high-voltage point. Total gap-length is 125 cm.)



IG. 1. Positive point electrode: earthed plane electrode. (*a*) $R = 100,000$: (i) $C = 10$; (ii) $C = 200$; (iii) $C = 5000$. (*b*) $C = 10$: (i) $R = 10,000$; (ii) $R = 100,000$; (iii) $R = 1,000,00$.

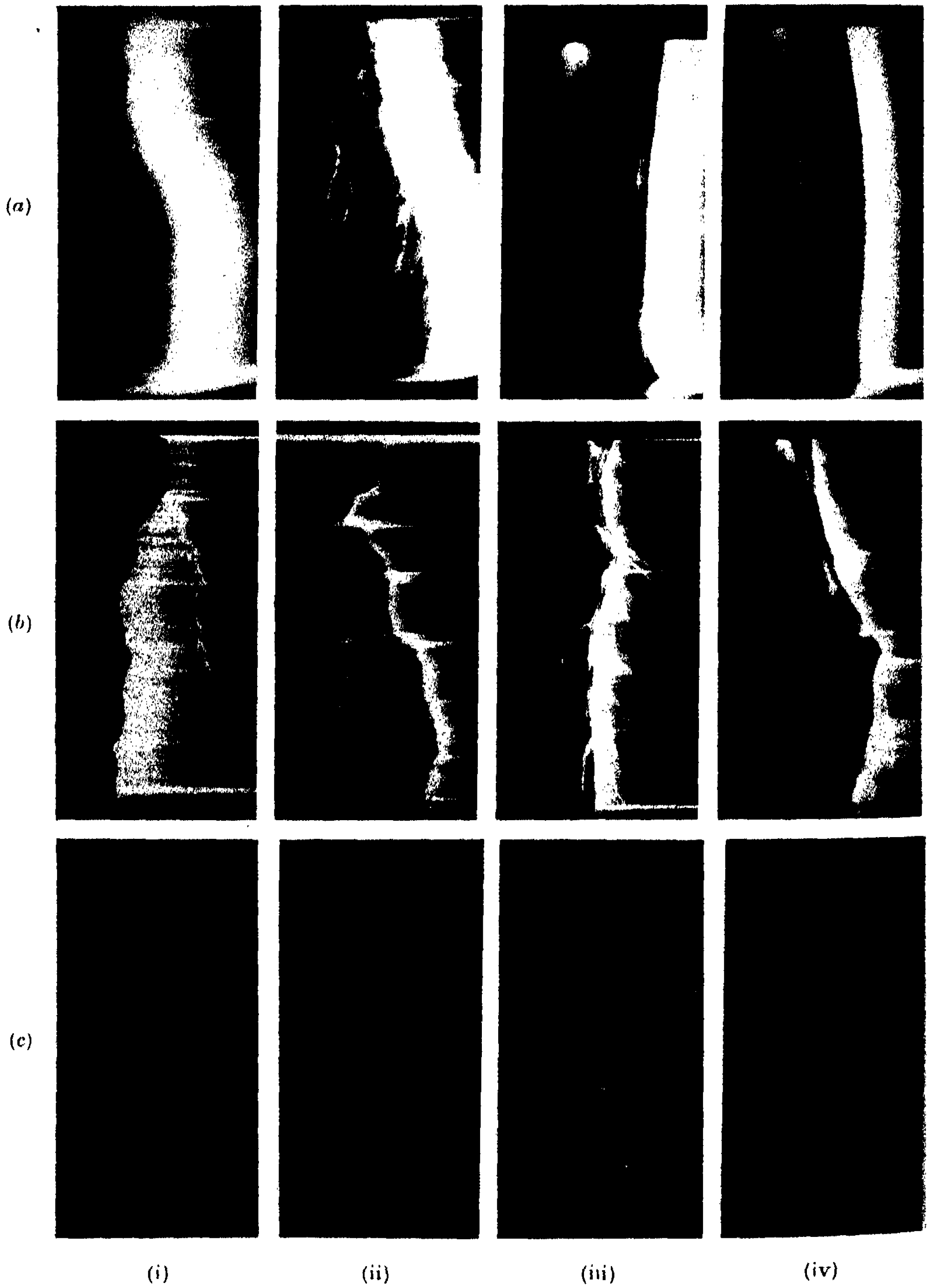


FIG. 3. Positive point electrode: earthed plane ($p = 1$ cm.) electrode. (*a*) $R = 700$; (*b*) $R = 10{,}000$; (*c*) $R = 100{,}000$. (i) $P = 40$; (ii) $P = 20$; (iii) $P = 10$; (iv) $P = 5$.

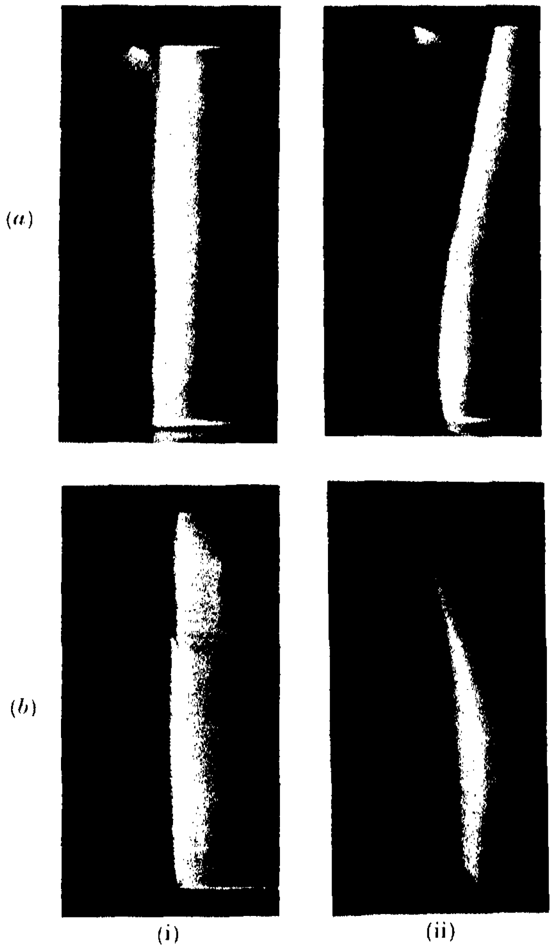
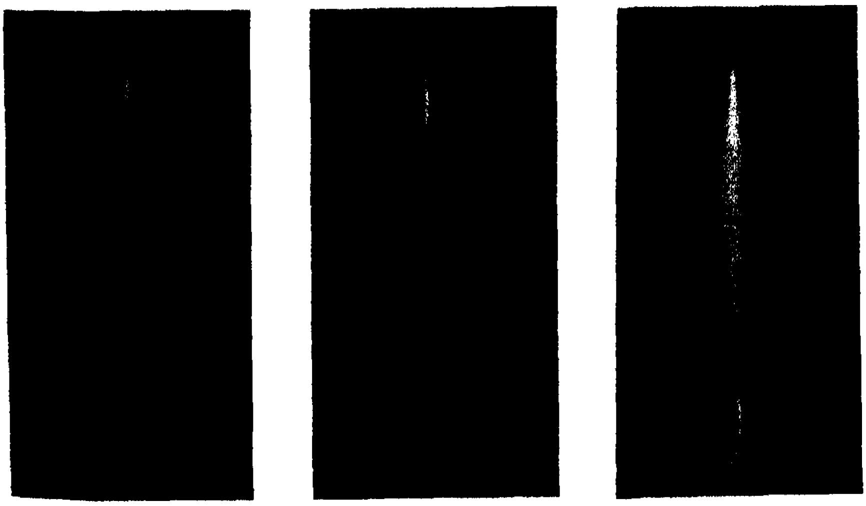


Fig. 4. Positive point electrode: earthed plane ($p = 1$ cm.) electrode. (*a*) $R = 700$; (*b*) $R = 10{,}000$. (i) $P = 2{\cdot}5$; (ii) $P = 1{\cdot}0$.



(*a*) (*b*) (*c*)

Fig. 5. **Positive point electrode: earthed plane ($p = 1$ cm.) electrode. "Still" photographs. $R = 700$. (*a*) $P = 0{\cdot}35$; (*b*) $P = 0{\cdot}6$; (*c*) $P = 0{\cdot}9$.**

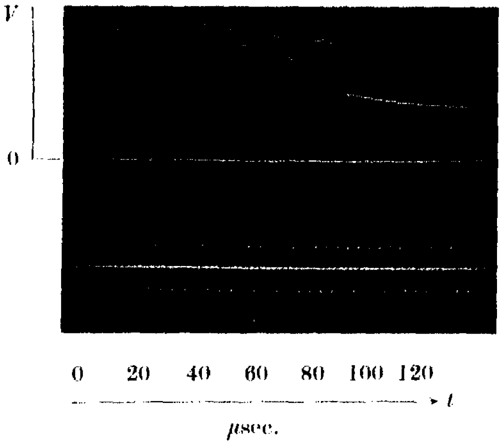
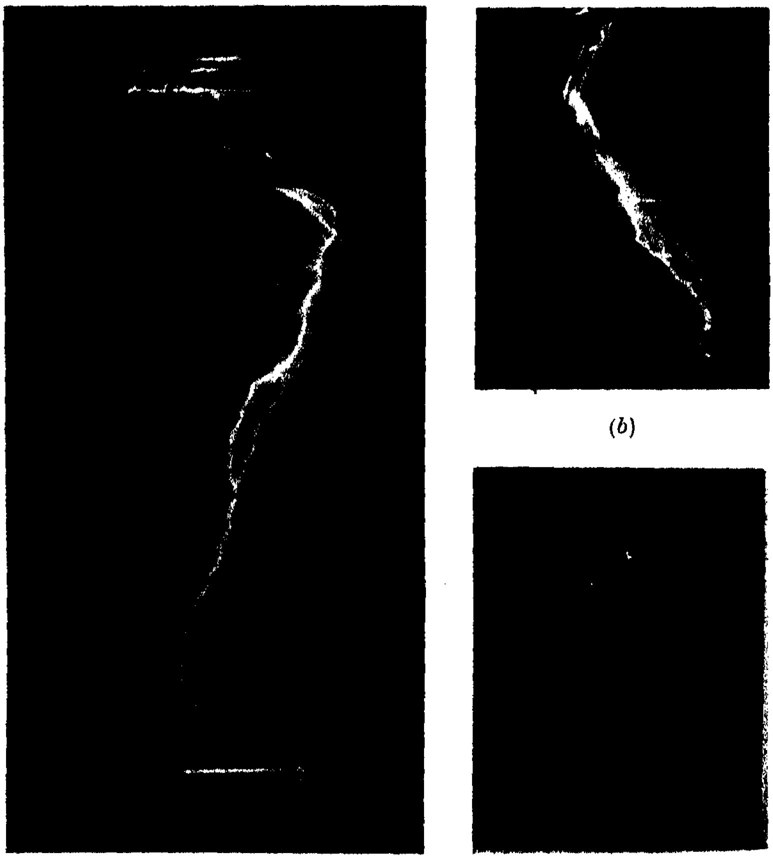


FIG. 8. Positive point electrode: earthed plane ($p = 1$ cm.) electrode. Voltage oscillogram. $R = 100,000$; $P = 10$. (Timing oscillation = 1 cycle per 4 μsec.)



(*b*)

(*a*) (*c*)

FIG. 13. Negative point electrode: earthed plane electrode. (*a*) $d = 100$; $R = 600,000$. (*b*) $d = 50$; $R = 600,000$. (*c*) $d = 100$; $R = 450,000$ (corona discharge shown for 40 % gap).

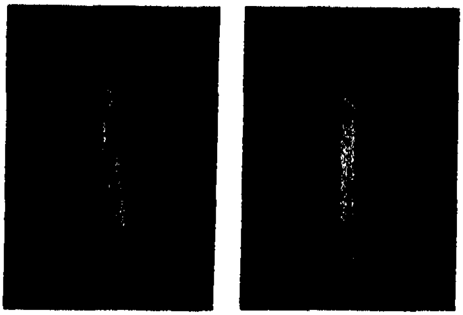

(*a*) (*b*)

FIG. 15. Negative point electrode: earthed plane electrode.
(*a*) $d = 25$; $R = 600{,}000$. (*b*) $d = 25$; $R = 2{,}000{,}000$.

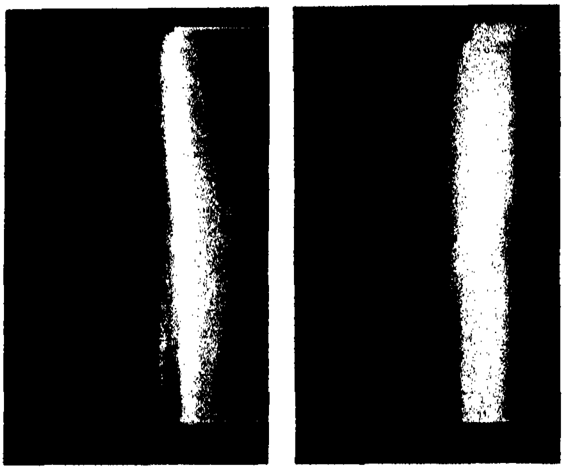

(*a*) (*b*)

FIG. 17. Negative point electrode: earthed plane ($p = 1$ cm.) electrode.
(*a*) $R = 10{,}000$; $P = 2{\cdot}5$. (*b*) $R = 700$; $P = 1$.

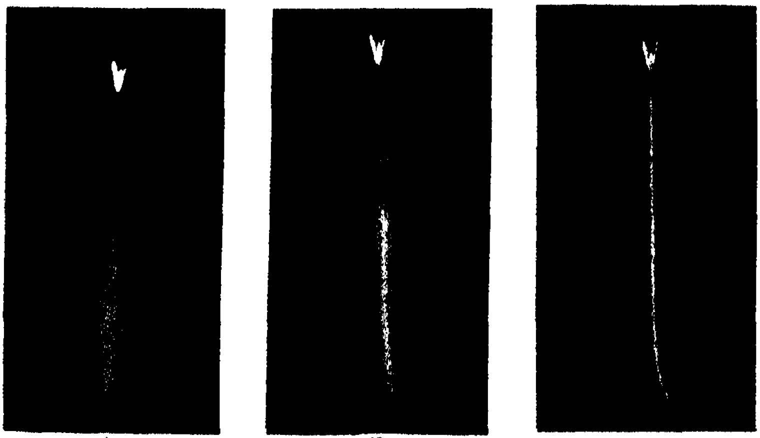

(*a*) (*b*) (*c*)

FIG. 18. Negative point electrode: earthed plane ($p = 1$ cm.) electrode.
"Still" photographs. $P = 0{\cdot}35$.

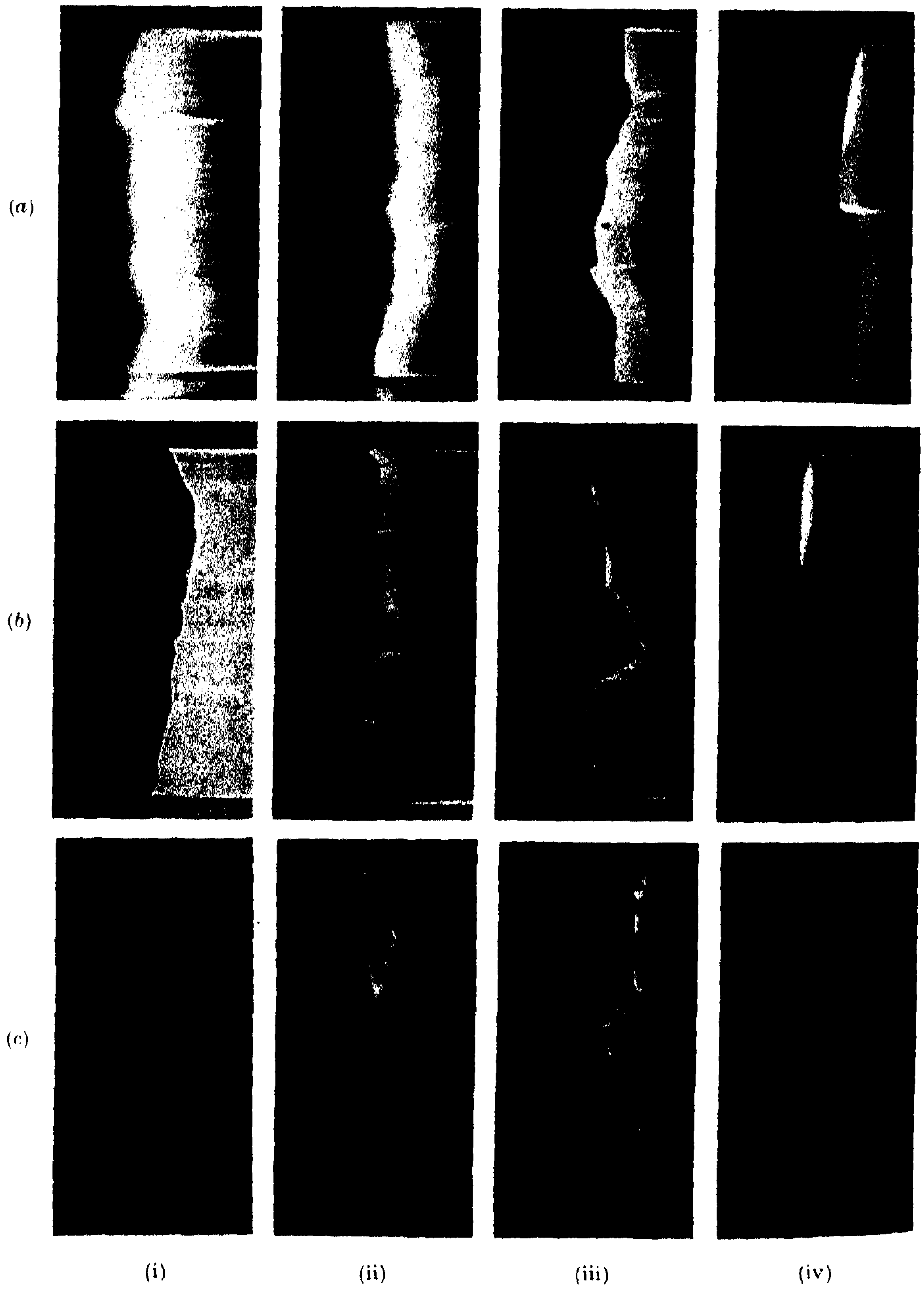


FIG. 16. Negative point electrode: earthed plane ($p = 1$ cm.) electrode. (*a*) $R = 700$; (*b*) $R = 10,000$; (*c*) $R = 100,000$. (*i*) $P = 30$; (ii) $P = 20$; (iii) $P = 10$; (iv) $P = 5$.

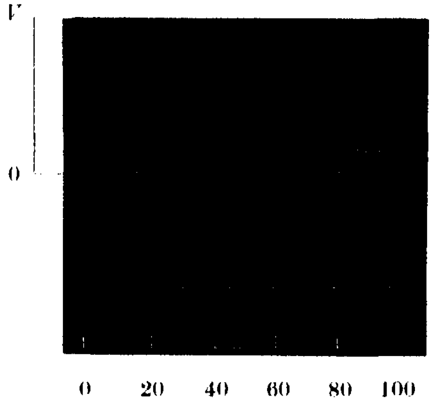


Fig. 20. Negative point electrode: earthed plane ($p = 1$ cm.) electrode. Voltage oscillogram. $R = 100,000$; $P = 10$. (Timing oscillation—1 cycle per 4 μsec.)

Fig. 22. Negative point electrode: earthed plane ($p = 1$ cm.) electrode. Current oscillogram. $R = 100,000$; $P = 19$. (Timing oscillation—1 cycle per 8 μsec.)

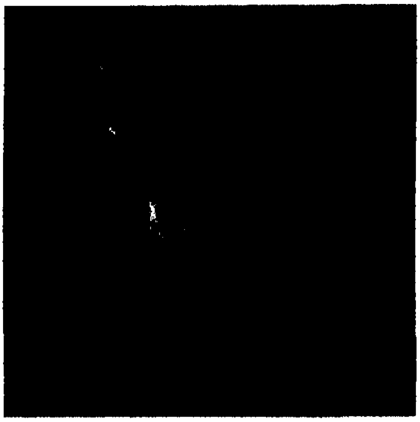


Fig. 23. Positive point electrode: earthed plane electrode. $R = 100,000$; $P = 5$.

(*b*) *Subatmospheric pressures*

The negative discharge has been investigated at pressures from 76 to 0·3 cm. Hg. Representative records for three different resistances are grouped in fig. 16, Plate 15: *a*, $R = 700$ ohms; *b*, $R = 10{,}000$ ohms; *c*, $R = 100{,}000$ ohms; while each group is divided into four pressures: (i) 30 cm. Hg; (ii) 20 cm. Hg; (iii) 10 cm. Hg; (iv) 5 cm. Hg.

With low resistance, as in the case of the positive discharge, leader strokes cannot be distinguished at atmospheric pressure due to halation caused by the main stroke. However, they are clearly visible at pressures below 30 cm. Hg, when both positive and negative leaders can be seen to meet in the upper mid-gap region. The height of the junction point increases with decrease in pressure. With the two higher resistances the stepped development of leader strokes is well defined, but there appears to be no fundamental difference in the mechanism of the discharge for all three pressures. The method of development of the leader strokes is most clearly seen in fig. 16 (*c*) (ii), Plate 15. Four brush discharges at the cathode are accompanied by four upgrowing positive leader strokes from the anode. No difference can be observed in the times of occurrence of these positive leader strokes and the corresponding negative discharges. A camera of higher resolving power might reveal such a difference, but at present it cannot be stated whether the positive leader is a consequence of the brush discharge appearing at the cathode, or whether the latter is produced as a result of the field distortion caused by the development of the positive leader. The last of the negative brush discharges takes the form of a leader stroke some 2 cm. in length: the lengths of the positive leaders are approximately 11, 16, 32 and 40 cm.—no continuous streamer can be observed to join their tips. Each of the positive leader strokes follows the track ionized by the previous leader, after which it develops into virgin air; each leader is strongly branched at its tip. The discontinuous nature of the development—seen at 10 cm. pressure—is not so apparent when the pressure is raised to 40 cm. Here the successive steps are so close that their tips appear to form a continuous streamer, but with a camera of higher resolving power it is probable that a discontinuous development would be observed.

At 5 cm. pressure the stepped development of the positive leader stroke becomes less apparent for all three resistances; however, in all cases the first long brush discharge is seen to accompany a short brush discharge at the cathode. An elongated column of discharge, apparently the junction-point of the positive and negative leaders, can be seen near the cathode; this column becomes more pronounced at still lower pressures. At 2·5 cm.

Hg the discharge—shown in fig. 17 (*a*), Plate 14—becomes much more diffuse and at 1 cm. Hg, fig. 17 (*b*), Plate 14, the upward developing leader is almost identical with the positive leader shown in fig. 4 (*a*) (ii): its speed is about 2×10^6 cm./sec., and it is continuous and not stepped. It merges into a very elongated column of discharge from the cathode. The stationary camera records given in fig. 18, Plate 14, for a pressure of 0·34 cm. Hg and for a range of applied voltages show that as the voltage is increased the central bright core to the diffuse glow discharge increases in length from the anode upwards. There is intense luminosity around the cathode point, and there is a dark space adjacent to it. With further increase of voltage sparkover can take place, unlike the positive discharge at the corresponding pressure. Further study with the rotating camera will be necessary to elucidate the progress of the discharge at very low pressures.

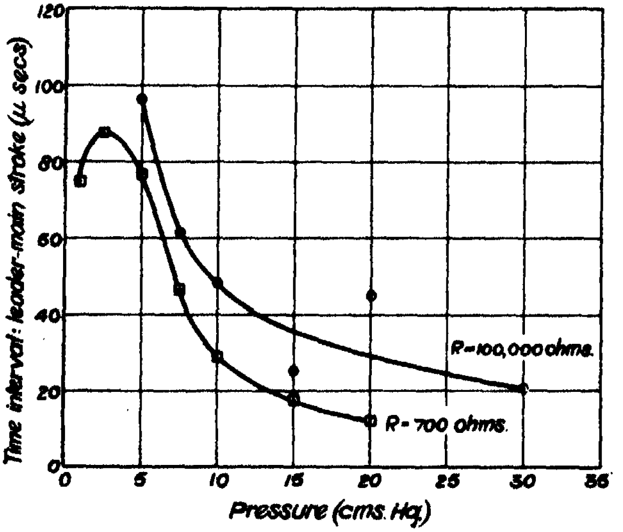


FIG. 19. Negative point electrode: earthed plane ($p = 1$ cm.) electrode. Curves relating the time interval between the initiation of the negative leader stroke and the main stroke with pressure for different series resistances.

The variation with pressure and resistance of the time interval between the initiation of the negative leader and the main stroke is shown in fig. 19, the curves are generally similar to those in fig. 6 for the positive discharge. At all pressures up to atmospheric the time intervals for the negative discharge are smaller than those for the positive discharge; a probable reason for this is that the negative discharge never traverses the whole gap but is met by an ascending positive leader from the earthed electrode.

A wide variation in voltage at the cathode occurs for high resistances when the pressure is low; a typical oscillogram of the voltage recorded on an antenna is shown in fig. 20, Plate 16. The variation in voltage is more pronounced than for the positive discharge, a fact which has previously been observed at atmospheric pressure (I). Calculations based on the measured voltage drop show that the current which flows on the occurrence of the second step in the oscillogram must be of the order of ½ amp. The

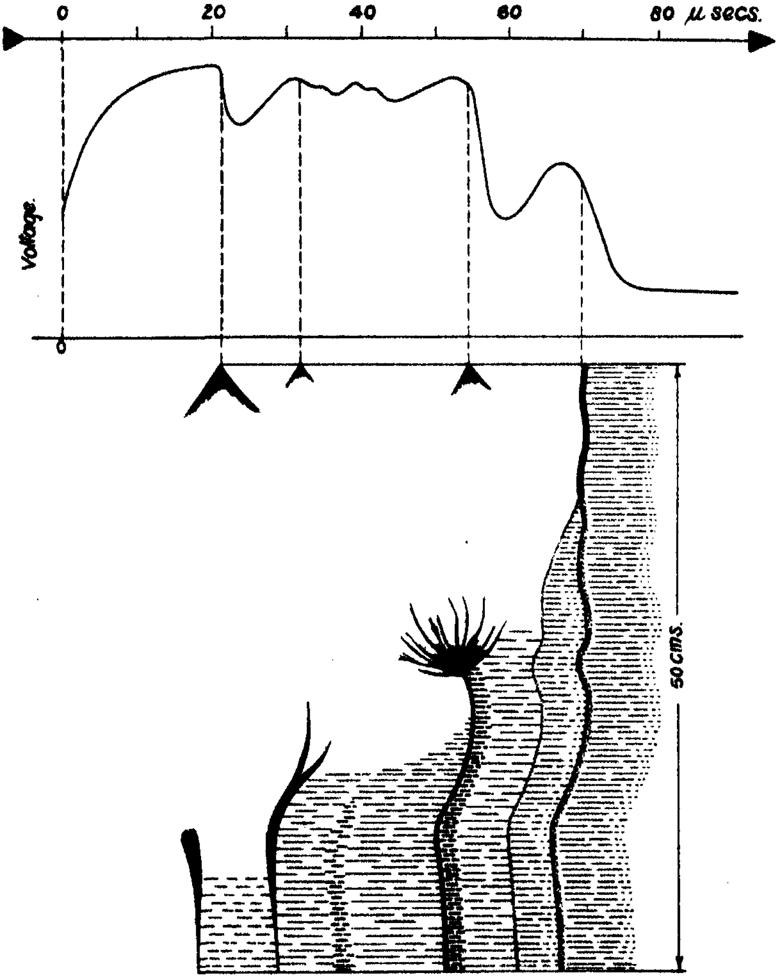


Fig. 21. Negative point electrode: earthed plane ($p = 1$ cm.) electrode. Drawing based on an original photograph of a discharge and the synchronized voltage oscillogram. $R = 100{,}000$; $P = 10$.

voltage fluctuations with series resistance of 10,000 and 1000 ohms are correspondingly smaller. An analysis of a photograph of a discharge together with the synchronized voltage oscillogram is given in fig. 21. The sudden drops in voltage are seen to correspond closely with the steps in the leader stroke development.

In the earlier paper (I) oscillograms of the current flowing to the earthed electrode were synchronized with photographs of the discharge in air at atmospheric pressure, and steps in the positive leader stroke were found to correspond to peak values of the current. The accentuation of the stepped nature of the positive leader development when the pressure is reduced causes the variation in the current to be more pronounced, and a current oscillogram at 18 cm. pressure is given in fig. 22, Plate 16. Subsequent to a period of 11 μsec. in which no current flows the leader stroke is initiated; the main stroke occurs after a further 35 μsec. during which time there are a number of peak values of current corresponding to the various steps in the leader stroke.

As in the case of the positive discharge a curve has been obtained to show the variation in the breakdown voltage of the discharge gap with pressure. Due to the frequent sparkover along the glass surface, particularly at higher pressures, the curve cannot be considered as accurate as that for the positive discharge. The curve—given in fig. 10—shows a similar deviation from Paschen's law to that observed for the positive discharge; this deviation may again be ascribed to the long time-lags to breakdown at the lower pressures.

6. Discussion of results

The further work reported in this paper on the development of the spark at atmospheric pressures confirms the observations given in (I) as regards the major differences between the positive and the negative spark discharge. Even for the longest gap (150 cm. breakdown voltage = 1600 kV) for which records are available the negative discharge from a point electrode is always met by an ascending positive leader stroke from the earthed plane and the length of this leader is not less than 30 % of the gap. In the case of the positive discharge the positive leader stroke is only infrequently met by an ascending negative leader stroke: the leader stroke of the positive discharge often splits into two main branches as it approaches the earthed plane but this branching is not necessarily caused by an ascending negative leader. The "root-branching" of the lightning flash referred to by Schonland and others (1935) and Schonland (1938) does not appear therefore to afford proof of the existence, before the main flash, of an upward-developing leader stroke.

The subatmospheric discharges afford assistance in the interpretation of the discharges at atmospheric pressure. The leader-stroke/main-stroke mechanism persists at least down to 5 cm. Hg with the same general distinction between positive and negative types of discharge. The individual stepped leader strokes become more sharply defined and more widely separated in time as the pressure is reduced (down to 10 cm. Hg). The positive leader strokes (whether downward or upward directed) at low pressures are seen to consist of a number of individual stepped leader strokes disconnected from one another; each stroke travels rapidly over a pre-ionized track and then more slowly through virgin air. A good example of this is given in fig. 23, Plate 16, where each step terminates as a fine tuft. At higher pressures the close proximity in time of the stepped leader strokes gives the impression of the pilot streamer referred to in (I) but, as stated previously, a camera of higher resolving power might show that the development of the discharge at atmospheric pressure was also discontinuous. The negative leader strokes at low pressures are shorter than at atmospheric pressure: at 10 cm. Hg they are little more than tufts of discharge 1 or 2 cm. in length disconnected from each other as at atmospheric pressure. At low pressures therefore the positive leader predominates to an even greater extent than at atmospheric pressure.

The mechanism of the spark has been investigated oscillographically by Rogowski, Flegler and Tamm (1927); Tamm (1928); Rogowski and Tamm (1928); Krug (1932) and Köhler (1936), using impulse voltages. Rogowski and others (1927) showed that the breakdown process between parallel plate electrodes took place in two stages, the voltage first collapsed from maximum value to about half that value, remained at about that value for a short time (some tenths of a microsecond) and then collapsed to zero. Tamm (1928) showed that this effect occurred at all pressures from 76 to 18 cm. Hg between parallel plate electrodes 1–4 mm. spacing. At very low pressures (0·7 cm.) the second stage of the breakdown process did not occur and it was noted that only a glow discharge took place, not a sharply defined spark. In air the step in the voltage collapse was not observed for breakdown in inhomogeneous fields. Rogowski and Tamm (1928) showed photographs of the glow discharge at various stages of development (at 4 cm. Hg between parallel plate electrodes) and they identified the start of the glow discharge with the first step in the voltage collapse: they noted that the next stage in the spark development started at the anode with the formation of a concentrated spark and this finally resulted in the complete voltage collapse. Krug (1932) observed the voltage step for breakdown in inhomogeneous fields, viz. between point/plane and sphere/plane electrodes. These experiments were made with electrode spacings of up to 10 cm., 80 kV. Köhler (1936) showed how the duration of the voltage step in the breakdown process varied with pressure and took oscillograms of the current in the discharge as well as of the voltage between electrodes: he obtained steps only with homogeneous fields, not with inhomogeneous fields.

No previous observations seem to have been made with the rotating camera in combination with the oscillograph, so that no definite correlation between the voltage steps and the development of the luminous discharge has been made. It appears from the present work that all the sudden discontinuities in the oscillograms of the voltage across the gap may be ascribed to the development of stepped leader strokes, i.e. the momentary voltage drop at the onset of each leader is the product of the circuit resistance and the leader-stroke current. Where the resistance in circuit is small the voltage steps may not be perceptible on the oscillograms: conversely they may be greatly accentuated by increasing the series resistance to high values as in fig. 8. The magnitude of the leader stroke currents appears to be of the order of one ampere in the present series of tests and also in the records of Rogowski and others where voltages of only 1–5 kV were used; to obtain good oscillograms of the step in the voltage collapse therefore the optimum value of circuit resistance should be increased linearly with the applied voltage.* The early predischarges carry smaller currents than the leader strokes which just precede the main stroke but they can be observed in the current oscillograms, and, with adequate series resistance, their occurrence can be identified in the voltage oscillograms. The duration of the voltage steps increases as the pressure is reduced to 5 cm. Hg, thereafter, for lower pressures, the duration diminishes as shown in figs. 6 and 19. The curves down to 5 cm. Hg are similar to those of Köhler (1936) for very small gaps and homogeneous fields. Below 5 cm. pressure the sharply defined positive and negative leader strokes become more diffuse, and at 1 cm. pressure the discharge corresponds to the glow discharge described by Tamm (1928). It is important to note that for large gaps, the discharge at low pressures develops from the positive electrode exactly as Rogowski and Tamm (1928) had described for very small gaps.

---

* In impulse voltage circuits it is customary to introduce sufficient resistance to make the circuits aperiodic, the value of this resistance is frequently of the order of some hundreds of ohms irrespective of the voltage of the circuit in which it is used. Whilst this value may be sufficient to accentuate the voltage steps on oscillograms of breakdown of very small gaps (as employed by Rogowski and others) a much higher value is in general necessary for large gaps.

Flegler (1937) has studied the development of the discharge at low pressures by means of the cloud chamber, and determined three main stages in the breakdown process: (*a*) preliminary stage of very weak illumination, generally only detectable as a fog in the chamber and sometimes by small irregularities of the voltage wave; (*b*) appearance of bright luminosity in a few channels, accompanied by the Rogowski "step" in the voltage wave; (*c*) complete sparkover and collapse of voltage. These three stages are confirmed in the present work, but the first stage is shown to be *essentially similar to the second stage, and differs from it only as* regards the magnitude of the current carried by the various predischarges.

The deviation from Paschen's law is not considered to be important. Köhler (1936) showed that Paschen's law held for power-frequency voltages and impulse voltages of a variety of waveshapes for pressures down to 10 cm. Hg, between spheres. The time to breakdown was small compared with the duration of the wave-tail of the impulse voltage. For point/plane electrodes the time to breakdown becomes so large at low pressures that impulse voltages higher than the static breakdown voltage have naturally to be applied.

Although the cause of the steps in the oscillogram of the breakdown process are now fully explained there is as yet no explanation of the variation of the duration of these with pressure. Since the critical breakdown field varies with pressure, and the mean free path varies inversely as pressure, the leader stroke velocity should not depend on pressure. A possible explanation for the slower velocity of the leader is that the density of photoelectrons ahead of the streamer diminishes with pressure, and, as Cravath and Loeb (1935) have shown, the velocity may depend on the density of primary electrons ahead of the streamer.

At low pressures the discharge develops almost entirely from the anode. A suggested explanation of this is that as the pressure is reduced the voltage gradient at the electrodes diminishes and, since the leader stroke current does not vary greatly with pressure, the cloud of positive ions formed at the cathode when a leader develops has a proportionately greater influence on the gradient ahead of this leader: this restricts the development of the negative leader as discussed in (I) (p. 125), and the completion of the breakdown process is left to the positive leader stroke.

The further study of the effect of resistance on the velocity of propagation of leader strokes reported here indicates the importance of this as a factor in discharge development. It is difficult to see how to make full use of this in studying the lightning flash but inasmuch as clouds have not the properties of metallic conductors it is reasonable to ascribe different leader-stroke velocities to differently charged clouds as previously suggested (Allibone 1938).

The authors with to express their thanks to Mr G. J. Scoles, M.Eng., for checking some of the oscillographic observations, and to Dr A. P. M. Fleming, C.B.E., Director and Manager of Research and Education Departments, Metropolitan-Vickers Electrical Co., Ltd., for permission to publish this paper.

REFERENCES

Allibone, T. E. 1938 *J. Instn Elect. Engrs.* **82**, 513.
Allibone, T. E. and Meek, J. M. 1938 *Proc. Roy. Soc.* A, **166**, 97. Referred to as (I).
Appleton, E. V. and Chapman, F. 1937 *Proc. Roy. Soc.* A, **158**, 1.
Cravath, A. and Loeb, L. B. 1935 *Physics*, **6**, 125.
Dunnington, F. G. 1931 *Phys. Rev.* **38**, 1535.
Flegler, E. 1937 *Elektrotech. Z.* **47**, 58.
Köhler, A. 1936 *Arch. Elect.* **30**, 528.
Krug, W. 1932 *Z. Tech. Phys.* **13**, 377.
Malan, D. J. and Collens, H. 1937 *Proc. Roy. Soc.* A, **162**, 175.
Rogowski, W., Flegler, E. and Tamm, R. 1927 *Arch. Elect.* **18**, 506.
Rogowski, W. and Tamm, R. 1928 *Arch. Elect.* **20**, 625.
Schonland, B. F. J. 1937 *Trans. S. Afr. Inst. Elect. Engrs.* **28**, 204.
— 1938 *Proc. Roy. Soc.* A, **164**, 132.
Schonland, B. F. J., Malan, D. J. and Collens, H. 1935 *Proc. Roy. Soc.* A, **152**, 595.
Tamm, R. 1928 *Arch. Elect.* **19**, 235.

The excitation of γ-radiation in processes of proton capture by light elements

By S. C. Curran, *St John's College*, P. I. Dee, *Sidney Sussex College*,
and V. Petržílka

(*Communicated by J. D. Cockcroft, F.R.S.—Received* 19 *September* 1938)

1. Introduction

In this paper the results of an investigation of the γ-radiations emitted in the process of proton capture by the four light elements, beryllium, boron, carbon and fluorine, are presented.

The intensities of the capture radiation for different energies of the incident protons have been determined from the numbers of coincidences produced in a pair of Geiger-Mueller counters. The high sensitivity and low natural effect of this arrangement render it very suitable for the detection of resonance effects of low intensity. Moreover, such weak resonances could be separated from more pronounced ones by the use of very thin targets from which the yield of γ-radiation was relatively low. By this method the excitation curves of the γ-ray emission have been determined up to an energy of the incident protons of 10^6 eV, and new resonances have been found for each of the above elements.

Measurements of the quantum energy of the emitted γ-radiation have been made for each of the four elements.

It has been possible, by analysis of the experimental data, to distinguish between the effects of proton capture by the constituent isotopes in the case of boron and of carbon.

In the case of carbon the yield of γ-quanta per incident proton has been found for both isotopes.

2. The experimental arrangement and detecting apparatus

The experiments here described were carried out in the Cavendish High Voltage Laboratory, using the Philips generator and the associated discharge tube installation from which beams of ions with energies up to $1{\cdot}2\times 10^6$ eV are available. Magnetic analysis of the accelerated ion beam was made, and the experiments are concerned only with the effects of proton bombardment.

The brass target tube *T* (fig. 1) contained a rotatable, water-cooled copper block *B* having two flat faces, either of which could be used to support targets in a plane perpendicular to the path of the beam. Usually, however, one of these two faces was used as a blank target for control experiments. The proton current incident upon the target ranged between 30 and 100 μA, being in most cases about 70 μA. The variation of the current in the course of an experiment was less than 2 %.

For the detection of the emitted γ-rays two thin-walled copper or brass Geiger-Mueller counters (C_1 and C_2) were employed in coincidence. The diameter of the tubes was 2 cm. and the distance of separation of their axes was 4 cm. The counters were placed inside a lead box with walls several cm. thick. An aperture *A* in this box permitted the passage of the capture γ-rays to the counters.

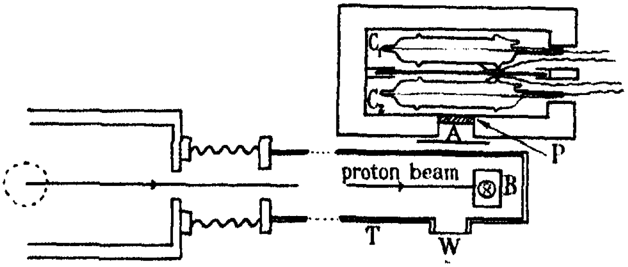


FIG. 1

A type of internally extinguished counter with thin walls was developed for use in these experiments. The thin-walled part of the tubes was about 6 cm. long and 0·1 mm. thick. The metal tubes were of copper or brass. The filling gas consisted of argon (partial pressure 9 cm.) and alcohol (partial pressure 1 cm.) (Trost 1935*a*, *b*, 1936, 1937). More details concerning these counters will be published elsewhere. These counters had a flat portion on their characteristic curves (counts × volts) of some 200 V.

A resistance of 1 MΩ was connected in series with each of the counters, and the voltage pulse developed across this resistance was applied to the grid of the first valve of a two-stage resistance-capacity coupled amplifier. The amplified pulses from the two counters were mixed in a Rossi (1932) circuit (fig. 2 *inset*), and coincidences were counted by a three-stage "scale-of-two" counting meter (Wynn Williams 1932; Lewis 1937).

Various factors influence the behaviour of the tubes in such a circuit. If N_1 counts/min. are produced in one counter and N_2 counts/min. are produced in the other counter, a number N_0/min. of chance coincidences will be recorded. N_0, N_1 and N_2 are connected by the equation $T = 60N_0/2N_1N_2$ sec., where T is known as the "resolving time" of the counting system. It was desirable to reduce the number of chance coincidences N_0 to a minimum. With our experimental arrangements there were two means of achieving this result. When the two counters were separated and screened from one another and two independent sources of β-radiation were used to irradiate them so that their individual rates of counting were N_1 and N_2/min., a number N_0/min. of chance coincidences was recorded when they were working in coincidence. For a fixed value of N_1N_2 it was found that N_0 increased rapidly with the voltage applied to the tubes as shown by curve 2 of fig. 2. If, however, the counters were placed close together and a source of β-radiation was placed so that the β-rays could pass through both tubes, producing real coincidences, the number of such real coincidences which was recorded for different voltages on the pair of tubes was found to be as shown in curve 1 of fig. 2. With a fixed strength of β-ray source the number of real coincidences per minute was constant over a large

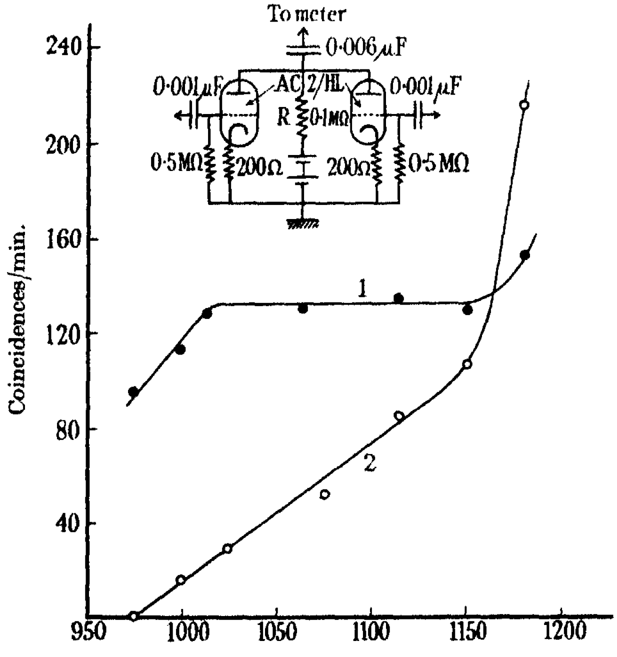


FIG. 2

range of voltage. In our experiments, therefore, the counters were operated at voltages which corresponded to the lower end of the flat portion of curve 1, fig. 2. It was also found that the resistance in series with each of the counters exerted a marked influence on the number of chance coincidences (N_0/min.) which was recorded for a fixed value of $N_1 N_2$. With our arrangement the optimum value of this resistance was found to be 1 MΩ. In addition to these two considerations a slight reduction in T could be obtained by suitable adjustment of discriminator bias. A resolving time of 10^{-5} sec. was finally obtainable with this simple and reliable arrangement.

We have used this counting system for the detection of the γ-radiation and for the measurements of its quantum energy. The capture radiation produced at the target B (fig. 1) and falling upon the aluminium plate P which closed the aperture A, ejected Compton electrons, photoelectrons and electron pairs from this plate. Some of these passed through both counters causing real coincidences. The range distribution of these electrons was determined by placing absorbing screens of different thicknesses between the counters and observing the number of coincidences as a function of the thickness of absorber. Information concerning the quantum energies of incident γ-radiations may be derived from such absorption curves in the following manner. The end-point of the absorption curve indicates the maximum range of the electrons and hence gives a measure of their limiting energy (Rasetti 1935). In the absence of precise data relating range and energy for electrons with energies of the order of 10^7 eV, it is necessary to obtain empirically a calibration curve connecting the observed end-point with the maximum quantum energy of the incident radiation. The curve which we have used for this purpose is shown in fig. 10, and an account of the method by which this curve has been obtained is given in § 5. Information concerning the quantum energy of the incident γ-radiation may also be derived from the value of absorber thickness at which the rate of coincidence counting falls to one-half of its value for no absorber between the counters (Fleischmann 1936). Mitchell and Langer (1937) and Kikuchi (1936*a*, *b*, *c*, 1937) have published data relating this half-value thickness with the quantum energy of the γ-radiation. It appears, however, that considerable divergence exists between the results of those who employ this method. A knowledge of this half-value thickness may often, however, be combined with a knowledge of the end-point to provide some information concerning the homogeneity or inhomogeneity of the incident radiation. For the study of radioactive effects produced by the capture of protons the target tube was provided with a window (W, fig. 1) of aluminium foil 0·1 mm. in thickness.

3. PREPARATION OF THE TARGETS

Carbon targets were obtained either by deposition of soot from a candle flame or by evaporation of aquadag. Both methods yielded targets of a uniform thickness which could be adjusted at will. The metallic beryllium used in the preparation of Be targets was an H.S. brand pure substance (obtained from Messrs Hilger, Ltd., Lab. No. 7778), spectroscopic examination of which had shown it to be free from lithium and fluorine. The beryllium targets were obtained as thin bright metallic layers upon copper plates by evaporation of this metal *in vacuo*. Very different thicknesses of target substance could be deposited in this manner. In the case of boron the same method was employed. Because the impurity content of the amorphous boron was not known, analytically pure B_2O_3 (Schering and Kahlbaum) was also used. No differences between the results for these two targets could be found. For the experiments with fluorine, BaF_2 (Schering and Kahlbaum) was evaporated on to the target block in air. Layers which were hardly visible were found to give sufficient intensity of γ-radiation for the purposes of the investigation and to remain unaltered during the course of an experiment.

4. EXPERIMENTAL RESULTS

(*a*) *Beryllium*

The excitation curve of the γ-radiation from beryllium under bombardment with protons has been examined by two groups of workers. Hafstad and Tuve (1935) first detected the capture of γ-radiation from this element using a thick target. Their results extended from about 600 kV (below which they could not detect radiation in sufficient intensity for accurate measurement) to 1000 kV, and they suggested that the process was not a resonance one. In later work, Herb, Kerst and McKibben (1937) extended the results beyond 1000 kV, and showed that a broad resonance was present at about 1 mV.

With our more sensitive detecting apparatus we have extended the study of the γ-radiation emitted in this process down to a voltage of 200 kV. We have examined, using pure Be targets of varying thicknesses, the course of the excitation curve from 200 to 1000 kV. The relationship between γ-ray intensity and bombarding voltage for a target of Be metal is shown in fig. 3, curve 1. The thickness of the target in this case is estimated as equal to the range of a proton of 20,000 eV energy. Our results are in rough agreement with those of Hafstad and Tuve for the region above 600 kV. Curve 2 of fig. 3 shows the results of a more detailed investigation of the region from

200 to 600 kV. It may be seen that a resonance peak is present at a bombarding voltage of 350 kV. The small peak at 480 kV and the rise of intensity at about 550 kV may be attributed to contamination of the target by carbon deposited during the bombardment (see § 4 (*d*)). A similar investigation of the region from 550 to 850 kV revealed a second new resonance with a peak intensity at 670 kV. This region is shown in more detail as curve 3 of fig. 3.

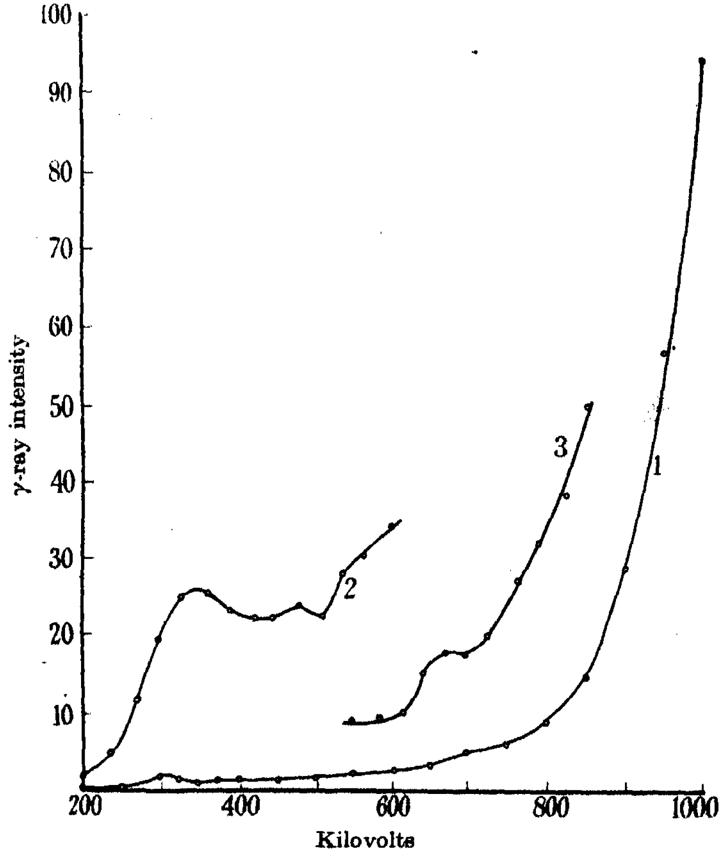


Fig. 3

The ion source used in these experiments was of the discharge tube type and an inhomogeneity of about 20,000 eV may exist in the energy of the protons from such a source. This inhomogeneity of the ion beam makes it impossible to draw any precise conclusions concerning the widths of the resonances which have been observed. It appears, however, from these results that the widths of the two new resonances at 350 and 670 kV, due to the bombardment of beryllium with protons, are considerably greater than the widths of the resonances associated with the bombardment of fluorine (see § 4 (*e*)).

The quantum energy of the capture radiation emitted in the process

$$^{9}\mathrm{Be} + {}^{1}\mathrm{H} \rightarrow {}^{10}\mathrm{B} + h\nu, \tag{1}$$

was investigated by Crane, Delsasso, Fowler and Lauritsen (1935*a*). Expansion chamber photographs of the tracks of electrons projected by the capture radiation gave evidence of the presence of quanta with energies 2·2, 3·7, 4·8 and 6·0 × 10^6 eV. For zero bombarding energy of the incident proton the maximum quantum energy available in process (1) is, by calculation from accepted mass values, 6·39 × 10^6 eV. The absorption curves which we have obtained for this element for different energies of the incident protons are shown in fig. 4. It may be seen from the results that the end points, according to our calibration curve, give values for the maximum quantum energy in good agreement with the values calculated from equation (1), when allowance is made for the different energies of the incident protons. These results prove that the radiation is, at least to a very great extent, due to the process of proton capture. In the case of beryllium bombarded with protons only one other process might lead to the emission of γ-radiation of appreciable energy:

$$^{9}\mathrm{Be} + {}^{1}\mathrm{H} \rightarrow {}^{6}\mathrm{Li} + {}^{4}\mathrm{He}. \tag{2}$$

A maximum quantum energy of about 2 × 10^6 eV would be possible from this reaction, but no excited states of ^{6}Li or of ^{4}He are known at such low energies. Moreover, the form of the absorption curves clearly shows that the bulk of the radiation is very much harder than 2 × 10^6 eV. Thus the half-value absorption thickness indicates, using the relationship due to Mitchell and Langer (1937), a γ-ray energy of 5·0 × 10^6 eV at 400 kV, and an energy of 5·7 × 10^6 eV at 850 kV. These results, while in accordance with the known inhomogeneity of the radiation, prove that most of the radiation arises from the process (1).

With the help of these facts concerning the observed energy release it is interesting to consider further the form of the excitation curve of the γ-radiation. It may be seen from the curve that the two resonances at 350 and 670 kV are weak effects superimposed upon a general background of radiation. The form of the curve is roughly exponential, the intensity of the background increasing smoothly with bombarding voltage, and it appears from the absorption measurements that this background radiation must be associated with the capture process (equation (1)). The excitation curve may then be regarded as due to the existence of a broad excited state in the resultant ^{10}B nucleus, extending between energy values of about 6·4 and > $7·4 \times 10^6$ eV above the ground state, the weak resonance effects at 350 and 670 kV being due to excited states of considerably smaller widths at energies 6·75 and $7·07 \times 10^6$ eV above the ground state.

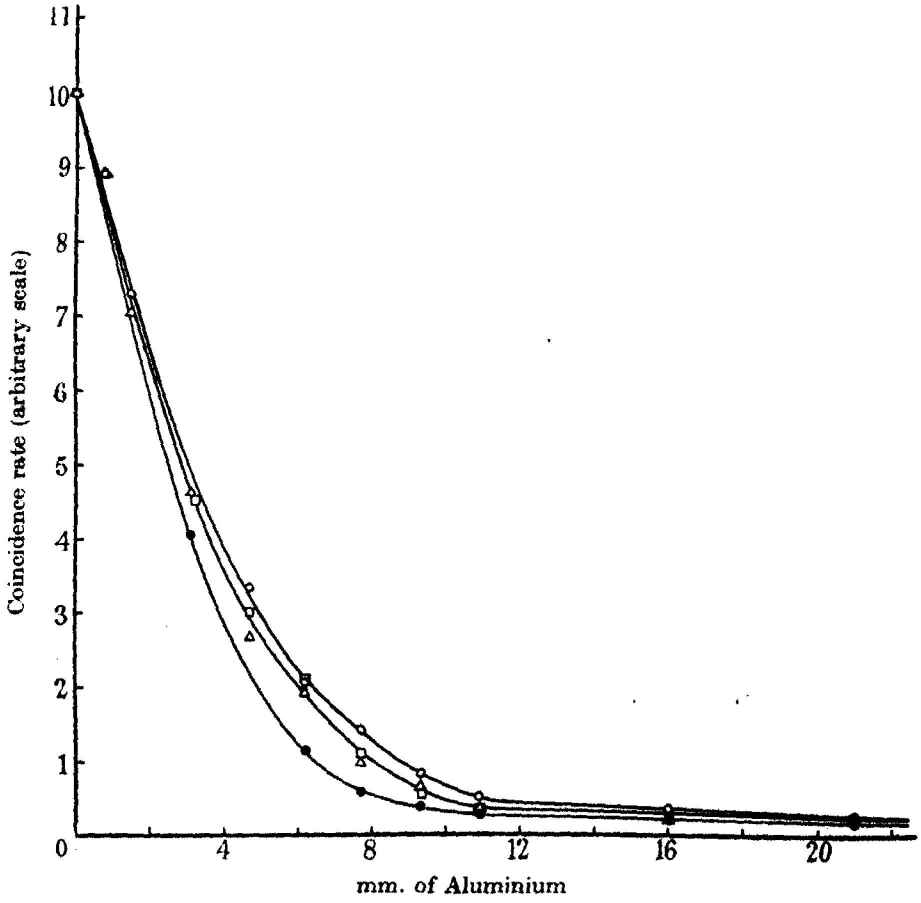


Fig. 4. Voltage symbols: ○, 850 kV; □, 700 kV; △, 600 kV; ●, 400 kV

		γ-ray energy from	
Voltage	End-point	End-point eV	Mass values eV
400 kV	11·1 mm.	$6·80 \times 10^6$	$6·75 \times 10^6$
650 kV	11·4 mm.	$7·00 \times 10^6$	$6·98 \times 10^6$
850 kV	11·7 mm.	$7·15 \times 10^6$	$7·16 \times 10^6$

In connexion with this possibility of the existence of a broad state of excitation of the ^{10}B nucleus it is interesting to notice that Gentner (1937) has found that the emission of γ-radiation from lithium under proton bombardment is not entirely confined to the regions of sharp resonance, a small but definite γ-ray emission being observed for all energies of the incident protons. It should be noted, however, that in the latter case the quantum energy is of the same magnitude as the energy release in the process of disintegration into α-particles, and the possibility of the broad resonance being due to the formation of α-particles in excited states has not therefore been excluded. We have given reasons for believing that the broad resonance observed in the case of beryllium cannot be attributed to excitation of the products of the reaction.

Alternatively, it is possible that the exponential form of the excitation curve is due to a lack of resolving power in the present experiments. It is possible, for example, that a large number of resonances exist in the region from 0 to 1 mV, the intensities of these resonances increasing with voltage and combining to produce an apparently exponential curve. On either view, the small changes of the maximum quantum energy which have been observed to be produced by changes of the bombarding voltage are to be expected.

(*b*) *Boron*

The excitation curve for the γ-rays emitted in the process

$$^{11}\mathrm{B} + {}^{1}\mathrm{H} \rightarrow {}^{12}\mathrm{C} + h\nu \qquad (3)$$

has been studied by Herb, Kerst and McKibben (1937), using a thick boron target of unknown purity. They obtained a smooth increase in intensity of γ-radiation with voltage, except for a weak resonance at about 820 kV. Bothe and Gentner (1937*a*) investigated the excitation at lower voltages from 100 to 500 kV. They found a resonance effect at 180 kV and a second rise at 360 kV, which they attributed to the effect of the molecular beam. They showed that the excitation curve of the γ-radiation ran parallel to the excitation curve of the α-particles produced in the process

$$^{11}\mathrm{B} + {}^{1}\mathrm{H} \rightarrow {}^{8}\mathrm{Be} + {}^{4}\mathrm{He}, \qquad (4)$$

which was studied by Williams and Wells (1936) and by Williams, Wells, Tate and Hill (1937).

Since it seemed desirable to obtain more accurate data on the excitation function between 300 kV and 1 mV we have investigated this region of voltage with great care. The results are shown in fig. 5. Curves 1 and 2 are the results obtained with thin targets of boron and boron trioxide.

The relative yields from these two targets may not be taken from the figure, since conditions of measurement were not the same in both cases. Curve 3 gives the measurements between 300 and 600 kV, using a B_2O_3 target, in greater detail. No evidence of any resonance effects may be derived from this curve. This confirms that the effect which Bothe and Gentner (1937*a*) found at 360 kV was due to the molecular beam as they suggested. It should be noted, however, that the γ-ray intensity as shown by this curve exhibits a fourfold increase between 300 and 500 kV, whereas Bothe and Gentner found an increase of only 70 % in this region. From 300 to about 800 kV the γ-ray intensity increases smoothly with voltage, but at 850 kV reaches a first resonance peak followed by a second one at 950 kV. There is also slight evidence of a change of gradient at about 650 kV, and this might indicate a third resonance.

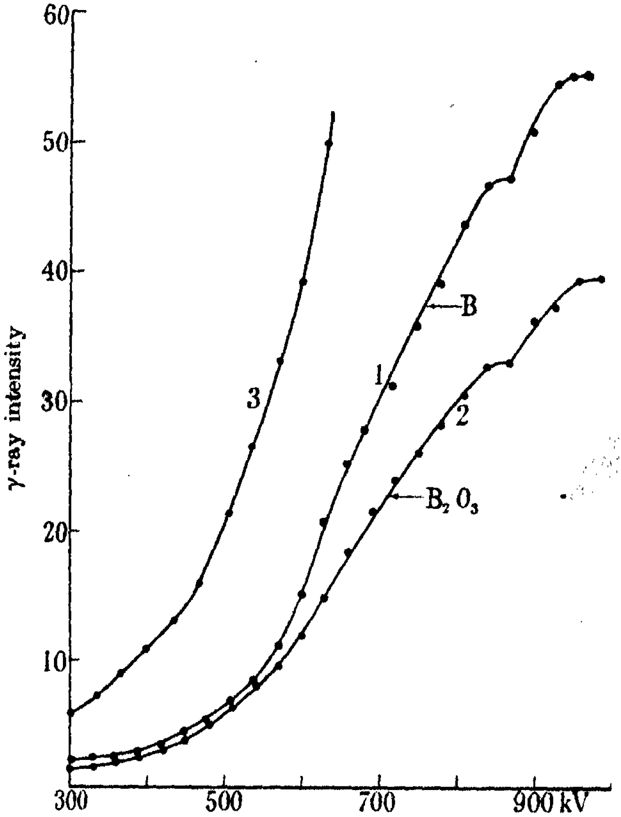


Fig. 5

Several energy measurements for the γ-rays emitted in process (3) have been carried out. The preliminary work of Crane, Delsasso, Fowler and Lauritsen (1935*b*) indicated a maximum quantum energy of $14{\cdot}5 \times 10^6$ eV. Bothe and Gentner (1937*a*) found a maximum quantum energy of 14×10^6 eV which was later re-determined by Gentner (1937) as 13×10^6 eV.

The maximum quantum energy to be expected by substitution of the mass values in process (3) is $16{\cdot}67 \times 10^6$ eV at a bombarding energy of 850 kV, and it was generally supposed that the resultant ^{12}C nucleus was left in an excited state of about 3×10^6 eV.

A detailed measurement of the γ-ray energy spectrum by Fowler, Gaerttner and Lauritsen (1938) has revealed the presence of three lines of energies 4·3, 11·8 and $16{\cdot}6 \times 10^6$ eV. It seems certain from these results that γ-rays with quantum energies corresponding to the maximum energy release indicated by equation (3) are emitted.

Measurements of the end-point of the absorption curve have given the value 28·0 mm. of aluminium at a bombarding voltage of 850 kV, and we have taken this value as corresponding to a quantum energy of $16{\cdot}67 \times 10^6$ eV in the construction of the curve relating absorption thickness with quantum energy (see § 5).

(c) *Radioactive effects from boron*

We have discussed above the capture process for boron without reference to capture by the isotope ^{10}B.

The formation of a radioactive positron emitter with half-life 20·5 min. has been reported by Crane and Lauritsen (1934*a*, 1935) to result from the bombardment of boron with protons having an energy of 900,000 eV. The formation of this radioactive body was attributed to the process

$$^{10}\mathrm{B} + {}^{1}\mathrm{H} \rightarrow {}^{11}\mathrm{C} + h\nu. \qquad (5)$$

The quantum energy of the γ-radiation to be expected from this process is $9{\cdot}4 \times 10^6$ eV, but no evidence of this energy value was found in the measurements of the γ-ray spectrum which were made by Fowler, Gaerttner and Lauritsen (1938). These latter workers considered therefore that the absence of any γ-rays of this energy might be explained by assuming that the production of radioactive ^{11}C was a resonance process occurring at a voltage of about 900 kV, a somewhat higher value than that used in their own experiments. Some support for this view appeared to be given by the fact that Allison (1936) found no evidence of the formation of a radioactive body with this period for proton energies as high as 500,000 eV.

In a series of experiments with proton energies varying from 350,000 to 750,000 eV we have been unable to find any definite evidence of the formation of a radioactive body with a half-life of 20·5 min. Since it appeared that in these experiments radioactive effects were being produced by the action of the γ-rays upon the target tube, etc., further experiments were made to eliminate such effects. A thick target of boron was bombarded with 60 μA of protons of energy 960,000 eV for 30 min., the target was withdrawn from the tube and the boron scraped away from the target holder. Analysis of the decay curve of the radioactivity of this boron showed the presence of two periods of values 10·7 min. and ~ 40 days. The most natural interpretation of this result is to associate the 10·7 min. period with radioactive ^{13}N formed by bombardment of ^{12}C present as contamination of the target, while the 40-day period may be attributed to radioactive ^{7}Be (half-period 43 days) formed as a result of the process $^{10}\mathrm{B} + {}^{1}\mathrm{H} \rightarrow {}^{7}\mathrm{Be} + {}^{4}\mathrm{He}$ (Roberts, Heydenburg and Locher 1938). Similar experiments carried out at different bombarding voltages yielded the same results. According to our measurements the yield of positron emission, having a half-life 20·5 min., is certainly less than 10^{-13} per proton, whereas the yield according to Crane and Lauritsen (1935) was stated to be about 10^{-10} to 10^{-11} per proton.

It is possible that the results of Crane and Lauritsen were due to photodisintegration effects, produced by the γ-rays from ^{11}B, since we found evidence of periods 10·5 and 38 min. in the earlier experiments in which the targets were not removed from the target tube. These periods could be attributed to the action of the γ-radiation upon the copper and zinc present in the brass of the target tube (Bothe and Gentner 1937*b*).

As a final test of the correctness of the above conclusions a boron target was bombarded with 50 μA of deuterons at 750 kV for 10 min. Analysis of the decay curve of the intense radioactivity which was produced in this case gave clear evidence of a period 21 min. which may be attributed to ^{11}C formed according to the process $^{10}\mathrm{B} + {}^{2}\mathrm{H} \rightarrow {}^{11}\mathrm{C} + {}^{1}_{0}n$. From the yield of the radioactive effect which was produced in this case it was clear that the amount of heavy hydrogen present in natural hydrogen would be sufficient to produce this activity in measurable amount. This observation may also be advanced as a possible explanation of the effect observed by Crane and Lauritsen who did not use magnetic separation of the bombarding beam.

(*d*) *Carbon*

The capture of protons by carbon according to the process

$$^{12}\mathrm{C} + {}^{1}\mathrm{H} \rightarrow {}^{13}\mathrm{N} + h\nu \qquad (6)$$

was investigated by Hafstad and Tuve (1935) who measured the amount of radioactive ^{13}N which was formed at different bombarding voltages. These workers detected the emission of γ-radiation but were unable to determine its excitation curve owing to the low sensitivity of their detecting apparatus. We have used counters in coincidence for the study of this capture radiation with the result shown in fig. 6, curve 1. Our earlier measurements (Dee, Curran and Petržílka 1938) have been extended to 950 kV. Two resonance effects at 480 and 570 kV are evident. We have also extended our earlier measurements of the positron emission from ^{13}N. This was done by bombarding the target for 3 min. at different voltages and measuring the subsequent activities. The fact that ^{13}N was formed was checked by evaluation of the period which was found to be 9·9 min. This agrees satisfactorily with the accepted value of the period making allowance for the escape of ^{13}N from the target, which was in vacuum during measurement. The results are shown in curve 2 of fig. 6. It is clear that only the lower resonance of curve 1 should be associated with the process (6). It is natural to associate the resonance at 570 kV with the process

$$^{13}C + {}^{1}H \rightarrow {}^{14}N + h\nu, \tag{7}$$

in which capture of protons leads to the formation of stable ^{14}N.

Consideration of the energy balance in these two reactions indicates that process (6) results in the release of γ-rays of quantum energy $2{\cdot}37 \times 10^6$ eV at a proton bombarding voltage of 470 kV, while for the process (7) the energy of the emitted γ-ray quantum is $8{\cdot}23 \times 10^6$ eV at 570 kV. Measurements of the γ-ray energies at 470 and 560 kV yielded the results shown in fig. 7. Our earlier work had indicated a γ-ray energy of $7{\cdot}4 \times 10^6$ eV at 560 kV, but further measurements with an improved counting system, at a bombarding voltage of 600 kV, indicated a γ-ray energy of $8{\cdot}5 \times 10^6$ eV (curve 1). This value of the quantum energy was obtained from the measured end-point value of curve 1. Consideration of the "half-value" absorber thickness indicates a quantum energy of about 6×10^6 eV. This fact may be explained by the assumption that the excited ^{14}N nucleus often reverts to the ground state by the successive emission of two quanta each of energy $\sim 4 \times 10^6$ eV (Rose 1938). This is in good agreement with the quantum energy $8{\cdot}2 \times 10^6$ eV expected from process (7). The energy of the γ-rays was also measured with a bombarding voltage of 480 kV (curve 2, fig. 7), the value of the quantum energy obtained in this case was $2{\cdot}6 \times 10^6$ eV. This is again in agreement with the quantum energy $2{\cdot}37 \times 10^6$ eV expected from process (6).

It has been shown therefore that between 400 and 900 kV there is one resonance (peak intensity at 480 kV) for the process (6), and between 300 and 950 kV one resonance (peak intensity at 570 kV) for process (7). A further confirmation of the above interpretation has been made in the

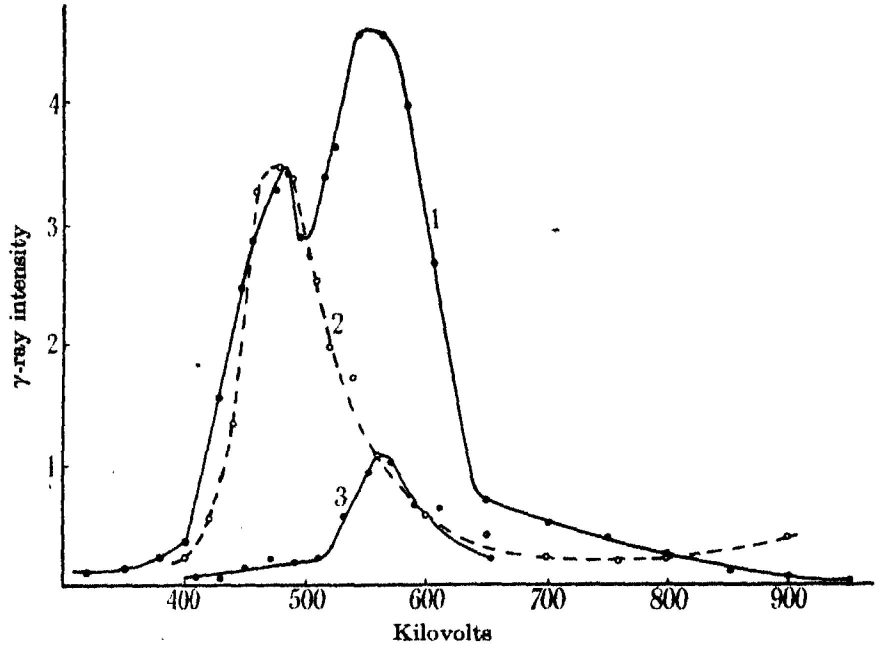

γ-ray intensity
1
2
3
4
1
2
3
400
500
600
700
800
900
Kilovolts

FIG. 6

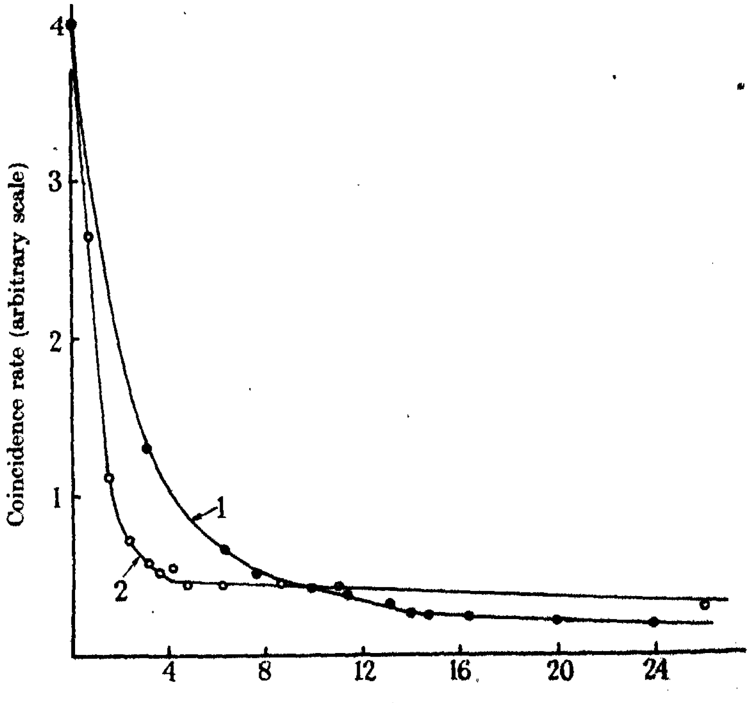

Coincidence rate (arbitrary scale)
1
2
3
4
1
2
4
8
12
16
20
24
mm. of Al.

FIG. 7

following manner. The excitation function of the emitted γ-radiation was measured with 6·2 mm. of aluminium between the counters (curve 3) fig. 6. The resonance at 480 kV was thus reduced by a factor 30, whereas the resonance at 570 kV was only reduced by a factor 4.

From the measured intensity of radioactive ^{13}N the yield of γ-ray quanta per incident proton in reaction (6) may be deduced. The total number of positrons emitted from a carbon target after bombardment with a known proton current for a known time interval was measured, correction being made for the absorption of electrons by the counter walls, decay of the radioactivity during bombardment, etc. The yield of positrons per incident proton was thus found to be $2{\cdot}5 \times 10^{-10}$. A recent measurement of this yield by Roberts and Heydenburg (1938) gave the value $3{\cdot}5 \times 10^{-10}$. Assuming that for each proton captured a single quantum of energy $2{\cdot}4 \times 10^6$ eV is emitted, the yield of quanta per incident proton is also, according to our measurements, $2{\cdot}5 \times 10^{-10}$. Some support for this assumption may be derived from the half-value absorption thickness (curve 2, fig. 7) which indicates a quantum energy of $2{\cdot}2 \times 10^6$ eV.

The yield of quanta per incident proton at 570 kV, in view of the fact that ^{13}C is about 1 % abundant in natural carbon, appears from the excitation curve 1, fig. 6, to be about 100 times greater than the value $2{\cdot}5 \times 10^{-10}$, which relates to capture by ^{12}C. Correction must be made, however, both for the fact that the proton capture by ^{13}C results in the formation of excited ^{14}N which often returns to the ground state by the emission of two quanta of energy 4×10^6 eV (Rose 1938), and for the different efficiencies of the counting system for the different energies of the γ-rays associated with the two resonances. With these corrections the γ-ray yield from process (7) is $6{\cdot}2 \times 10^{-9}$ quanta per incident proton. The ratio of the cross-sections for proton capture by ^{13}C and by ^{12}C is therefore approximately 25/1.

(*e*) *Fluorine*

McMillan (1934) and Crane and Lauritsen (1934*b*) first observed the γ-rays emitted in the process of proton capture by fluorine. The excitation function of the capture radiation has been measured by Hafstad and Tuve (1935). Hafstad, Heydenberg and Tuve (1936) and Herb, Kerst and McKibben (1937) have confirmed some of the earlier observations. The existence of sharp resonances at 330, 890 and 940 kV was established. Between 400 and 700 kV the form of the excitation curve suggested the presence of several weak resonances, but no conclusive results were obtained owing to the difficulty of preparing a suitable target. We have studied the excitation function for the capture process

$$^{19}F + {}^{1}H \rightarrow {}^{20}Ne + h\nu \tag{8}$$

for bombarding voltages between 300 and 1000 kV. The results obtained with very thin fluoride targets are shown in fig. 8. Curve 1 shows the results obtained in a run over the whole range of voltage, readings being taken in steps of 30 kV. It may be seen that confirmation of the earlier results has been obtained in the observation of resonance peaks at 340, 860 and 920 kV. The presence of two further peaks between 500 and 700 kV was also suggested. In view of the possible existence of other resonances in this region of voltage we have made a more detailed measurement of the

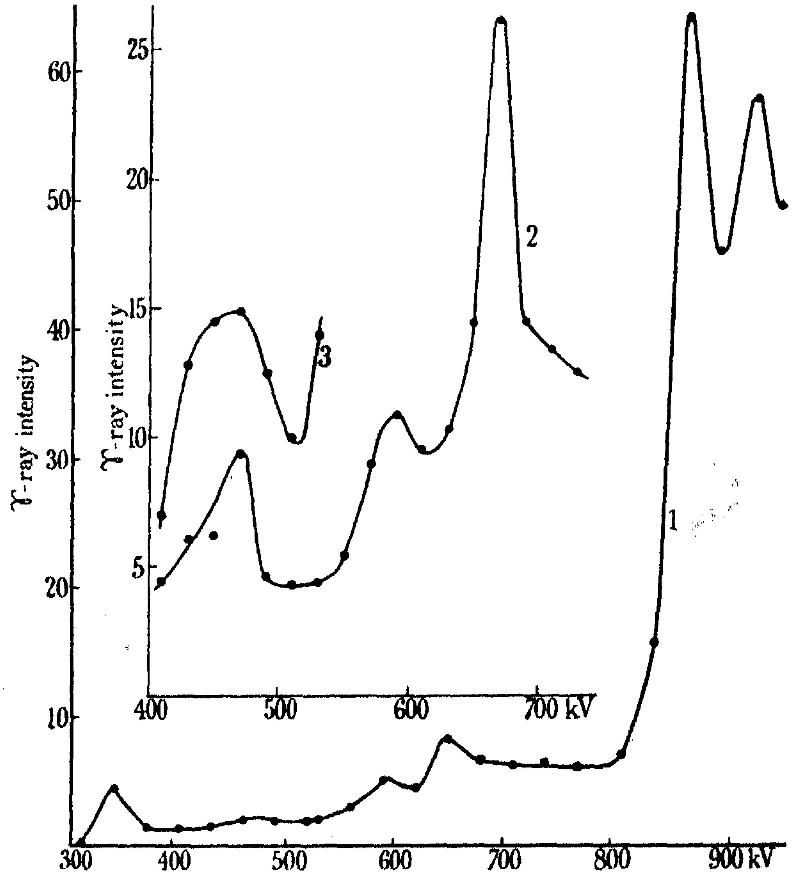


Fig. 8

excitation function between 400 and 750 kV. The results of these measurements are shown in curve 2, fig. 8. It may be seen that in this region there are three resonances with peaks at 470, 590 and 670 kV. The same three resonances have been found in several independent sets of observations with both thick and thin targets. In view of the possibility that the resonance at 470 kV may have been due to the ^{12}C resonance at 480 kV, this part of the excitation curve was remeasured with 4 mm. of aluminium placed between the counters. This thickness of absorber would have eliminated any effect due to the $2\cdot4 \times 10^6$ eV γ-radiation emitted in the ^{12}C resonance process. The result of this measurement is shown by curve 3, and it may be seen that no significant change in the excitation curve was produced. It may therefore be concluded that these three resonance peaks are to be associated with excitation levels of ^{20}Ne produced according to equation (8).

The absorption curve obtained (at 600 kV) in the measurement of the quantum energy of the radiation is given in fig. 9, curve 1. A more precise determination of the end-point is shown in curve 2. The statistical fluctuations for the points of curve 2 are too small to be entered in the figure. The end-point at 10·2 mm. of aluminium gives a quantum energy of $6\cdot3 \times 10^6$ eV. Quanta of the full energy available according to equation (8) ($13\cdot5 \times 10^6$ eV) are not emitted in this process. The half-value absorption thickness indicates a quantum energy of $5\cdot5 \times 10^6$ eV in agreement with the known inhomogeneity of the radiation.*

5. THE CALIBRATION CURVE

We have found in this work that with the counters described and for suitable intensities and types of γ-radiation the end-point of the electron absorption curve may be accurately determined and affords a satisfactory method of determining the upper limit of the γ-ray energy with an error of less than 5 %. With a pair of counters whose walls were 2 cm. apart it is possible to work accurately at counting rates of 1000 coincidences/min., and rates as high as 6000 coincidences/min. have been used. The component of the coincidence intensity remaining after the β-rays are absorbed lies between 1/20 and 1/50 of that with zero thickness of absorption. In the case of fluorine for example (fig. 9, curve 1) the initial intensity for zero thickness of absorber was 6000 coincidences/min. For curve 2 the rate of counting in the background itself was 1500 coincidences/min. The initial intensity in this case could not be measured on any of the counting meters available. The end-points obtained from these two measurements were almost identical. It is clear that there is no need to employ low rates of coincidence counting when sufficient intensity is available. It may be remarked that the time required in such cases to determine the γ-ray energy limit is about 1 hr. With softer radiations, however, having quantum energies of about 10^6 to 2×10^6 eV the end-point determination is rendered more difficult by the fact that many of the β-rays which are produced pass less readily through both counters.

---

* Unpublished results of Mr Bower indicate a value of $6\cdot3 \times 10^6$ eV for the maximum quantum energy liberated in this process.

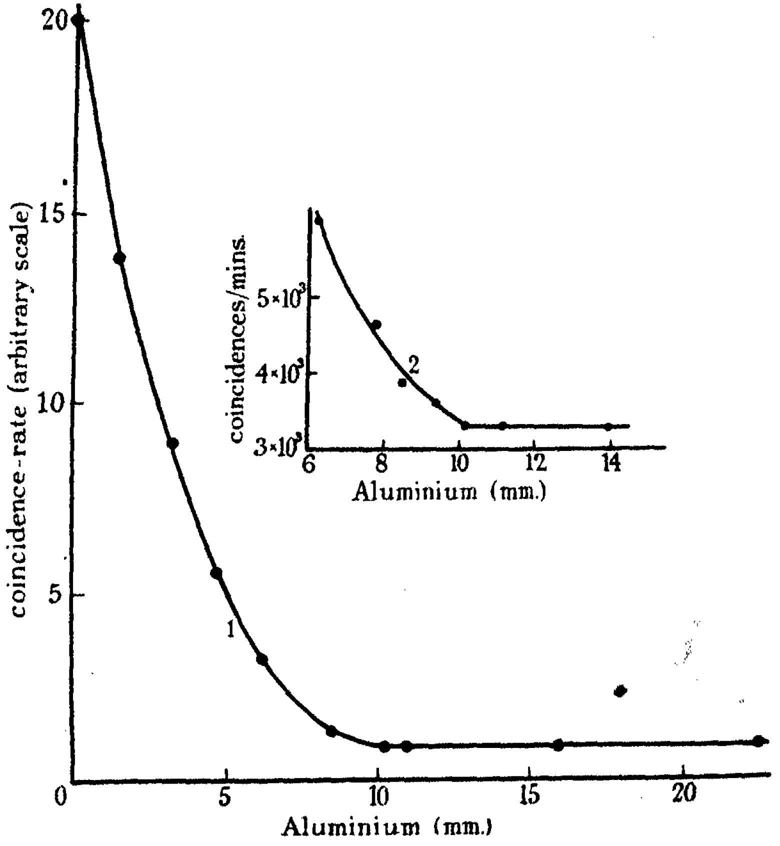


Fig. 9

For the conversion of the measured end-point absorption values into quantum energies a calibration curve (fig. 10) was constructed in the following manner. The results of Bernardini and Franchetti (1937) were plotted (open circles). The end-point for the γ-rays from radio-sodium, of quantum energy $3{\cdot}08 \times 10^6$ eV was measured and found to lie upon the straight line passing through these points. The measured end-point value (28·0 mm.) for the capture radiation from boron under bombardment with protons of energy 850,000 eV was taken to correspond with the value $16{\cdot}7 \times 10^6$ eV, which was found by Fowler, Gaerttner and Lauritsen (1938) for the quantum energy of the γ-radiation. It may be seen that this point lies upon the linear extrapolation of the points relating to the lower energies.

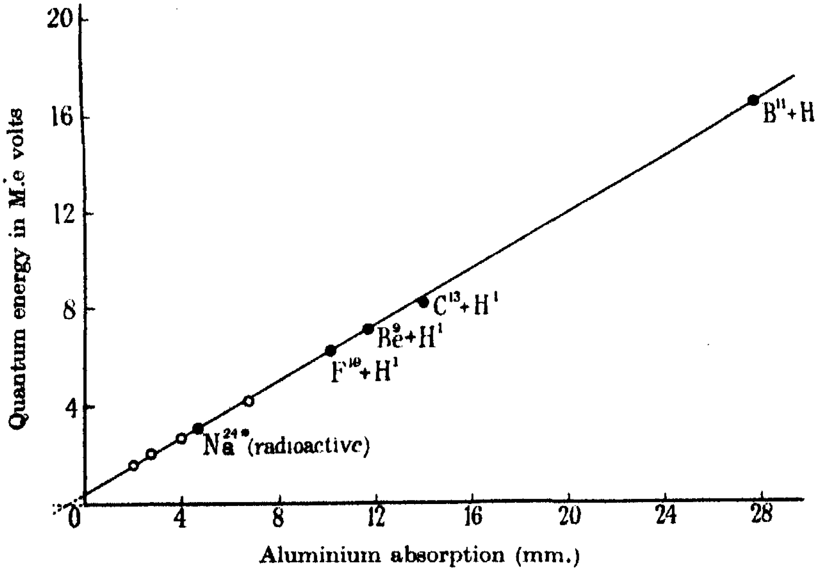


FIG. 10. ○, Bernardini and Franchetti. ●, Curran, Dee and Petržilka.

We have also plotted upon this diagram other points determined by our measurement of end-points and values of quantum energies calculated from known mass values. This calibration curve seems therefore to be essentially correct for quantum energies up to 17×10^6 eV. The line does not pass through the origin, since the absorption of the counter walls was equivalent to about 0·6 mm. of aluminium.

SUMMARY

Using a pair of Geiger-Mueller counters in coincidence the excitation functions for the γ-radiations emitted in the process of proton capture by each of the elements, beryllium, boron, carbon and fluorine have been determined for energies of the incident protons from 0 to 10^6 eV.

Resonance peaks have been found in the excitation curves for incident protons of the energies given below:

Beryllium	350, 670 × 10^3 eV
Boron	180, 650, 850, 950 × 10^3 eV
Carbon: ^{12}C	480 × 10^3 eV
^{13}C	570 × 10^3 eV
Fluorine	330, 470, 590, 670, 860, 920 × 10^3 eV.

Several of these resonance peaks had not hitherto been discovered.

In the cases of beryllium and boron a continuous excitation of γ-radiation over the whole of the region of voltage employed was observed, and evidence is presented in support of the view that this radiation must be associated with the capture process.

The yield of γ-radiation was measured in the case of carbon, the values obtained being $2 \cdot 5 \times 10^{-10}$ quanta/proton for ^{12}C and $6 \cdot 2 \times 10^{-9}$ quanta/proton for ^{13}C.

A calibration curve is given relating the end-points of the coincidence absorption curves (obtained by placing aluminium between the counters) with the maximum quantum energies of the incident γ-radiations.

A search for the production of radioactive ^{11}C as a result of the capture of protons by boron has shown that the cross-section for this effect is at least 100 times smaller than the value which has been reported by other workers.

REFERENCES

Allison 1936 *Proc. Camb. Phil. Soc.* **32**, 179.
Bernardini and Franchetti 1937 *Ric. sci.* **8**, 406.
Bothe and Gentner 1937*a* *Z. Phys.* **104**, 685.
— — 1937*b* *Z. Phys.* **106**, 236.
Crane, Delsasso, Fowler and Lauritsen 1935*a* *Phys. Rev.* **47**, 782.
— — — — 1935*b* *Phys. Rev.* **48**, 102.
Crane and Lauritsen 1934*a* *Phys. Rev.* **45**, 497.
— — 1934*b* *Phys. Rev.* **46**, 531.
— — 1935 *Phys. Rev.* **48**, 102.
Dee, Curran and Petržilka 1938 *Nature, Lond.*, **141**, 642.
Fleischmann 1936 *Z. Phys.* **103**, 113.
Fowler, Gaerttner and Lauritsen 1938 *Phys. Rev.* **53**, 628.
Gentner 1937 *Z. Phys.* **107**, 354.
Hafstad and Tuve 1935 *Phys. Rev.* **48**, 306.
Hafstad, Heydenburg and Tuve 1936 *Phys. Rev.* **50**, 504.
Herb, Kerst and McKibben 1937 *Phys. Rev.* **51**, 691.
Kikuchi and collab. 1936*a* *Nature, Lond.*, **137**, 992.
— 1936*b* *Nature, Lond.*, **138**, 841.
— 1936*c* *Proc. Phys. Math. Soc. Japan*, **18**, 727.
— 1937 *Proc. Phys. Math. Soc. Japan*, **19**, 734.

Lewis 1937 *Proc. Camb. Phil. Soc.* **33**, 549.
McMillan 1934 *Phys. Rev.* **46**, 325.
Mitchell and Langer 1937 *Phys. Rev.* **52**, 137.
Rasetti 1935 *Z. Phys.* **97**, 64.
Roberts and Heydenburg 1938 *Phys. Rev.* **53**, 374.
Roberts, Heydenburg and Locher 1938 *Phys. Rev.* **53**, 1016.
Rose 1938 *Phys. Rev.* **53**, 844.
Rossi 1932 *Nature, Lond.*, **125**, 636.
Trost 1935*a* *Phys. Z.* **36**, 801.
— 1935*b* *Z. tech. Phys.* **16**, 457.
— 1936 *Z. Phys.* **100**, 553.
— 1937 *Z. Phys.* **105**, 399.
Williams and Wells 1936 *Phys. Rev.* **50**, 187.
Williams, Wells, Tate and Hill 1937 *Phys. Rev.* **51**, 434.
Wynn Williams 1932 *Proc. Roy. Soc.* A, **136**, 312.

Lewis 1937 *Proc. Camb. Phil. Soc.* **33**, 549.
McMillan 1934 *Phys. Rev.* **46**, 325.
Mitchell and Langer 1937 *Phys. Rev.* **52**, 137.
Rasetti 1935 *Z. Phys.* **97**, 64.
Roberts and Heydenburg 1938 *Phys. Rev.* **53**, 374.
Roberts, Heydenburg and Locher 1938 *Phys. Rev.* **53**, 1016.
Rose 1938 *Phys. Rev.* **53**, 844.
Rossi 1932 *Nature, Lond.*, **125**, 636.
Trost 1935*a* *Phys. Z.* **36**, 801.
— 1935*b* *Z. tech. Phys.* **16**, 457.
— 1936 *Z. Phys.* **100**, 553.
— 1937 *Z. Phys.* **105**, 399.
Williams and Wells 1936 *Phys. Rev.* **50**, 187.
Williams, Wells, Tate and Hill 1937 *Phys. Rev.* **51**, 434.
Wynn Williams 1932 *Proc. Roy. Soc.* A, **136**, 312.

Relaxation methods applied to engineering problems
IV. Problems relating to elastic stability and vibrations

By K. N. E. Bradfield, D.Phil., D. G. Christopherson, B.A., and R. V. Southwell, F.R.S.

(*Received* 3 *August* 1938)

Introduction

1. Preceding papers in this series have been concerned with problems of a kind in which the solution (e.g. the configuration assumed by an elastic system in static equilibrium) is *unique*. Here we extend the Relaxation Method to problems such that a complete solution would define an infinite number of configurations, all satisfying the governing equations and the terminal conditions, but each associated with some particular value of a critical loading or frequency (the "characteristic number" or *Eigenwert*).

In Part II the convergence of the method as applied to problems of elastic equilibrium was established by the argument that every permitted "operation" reduces the total potential energy, which quantity has its minimum value in the configuration sought. By a corresponding argument its convergence can be established in relation to other problems having unique solutions; but in problems of the kind considered in this paper, although the "characteristic numbers" are stationary values of a particular (integral) function of the co-ordinates, only the lowest of them is a ***minimum*** value, and this circumstance necessitates a fundamental revision of the relaxation technique.

2. Two problems are treated in detail—namely, the calculation (1) of critical ("Euler") values of end-thrust and (2) of natural frequencies of transverse vibration for a straight rod of non-uniform mass and flexural rigidity. Some attention was given to the first problem in a paper, published elsewhere (Atkinson, Bradfield and Southwell 1937), which extended the two alternative methods of Part I to cover the case of flexure due to transverse loads and end-thrust acting in combination: it was shown that the first method gives (when extended) a means of determining the critical loads of a strut. In this paper the second method (more properly styled a relaxation process) is shown to lead more quickly to the same result. Constraints are imposed at selected cross-sections, and transverse displacements are permitted: account is then kept of the forces imposed on constraints by reason (i) of flexural rigidity and (ii) of end-thrust, the objective being to balance these forces at every section by modifying (*a*) the mode and (*b*) the magnitude of the thrust. The modification is effected "by inspection": that is to say, at every stage the distribution of the "residual forces" is examined and from it the next (tentative) operation is decided. In §§ 8–11 a form of strut is treated for which the exact solution has been obtained (by Cowley and Levy 1918): the first four critical loads are determined with percentage errors 0·09, 0·5, 1·5 and 2·9 respectively, by a relaxation process using 10 points of subdivision; when 20 divisions are taken, the error in the fourth critical load is reduced to 0·2 %. Compared with Rayleigh's method (Temple and Bickley 1933, § 3·1) the relaxation method has the important merit of determining with precision not only the critical loads but the mode appropriate to each.

3. For the strut a relaxation "by inspection" leads quickly to the required solution, since there is never any doubt as to what should be the next step. In comparison the problem of whirling or of transverse vibrations—in which both the mass and flexural rigidity influence the frequencies, and either may have any specified distribution—was found to proceed much more slowly when attacked in this manner: it was sometimes very difficult to decide what should be the next operation. Accordingly we propose for such problems (§§ 25–28) a systematic procedure based on Rayleigh's theorem regarding the stationary property of the natural frequencies. Approximate formulae (strictly defensible only when the mode is nearly correct, but improving in accuracy at every operation) are applied in the

normal manner to determine, successively, the first, second, third, etc., mode and natural frequency. In §§ 13–21 similar methods are propounded for use in connexion with struts if required. As stated above they will only exceptionally be needed, but they form a convenient introduction to the harder vibration problem.

PART I. CRITICAL LOADS, AND THE CORRESPONDING MODES OF DISTORTION, FOR A STRUT OF NON-UNIFORM SECTION

THE GOVERNING EQUATION

4. The governing equation of a strut, initially straight, which is held deflected by end-thrust unaccompanied by transverse loading is*

$$B\frac{d^2y}{dx^2}+Py=0, \tag{1}$$

where B stands for the flexural rigidity at a section defined by x,
P stands for the end-thrust, and
y stands for the deflexion away from the line of thrust.

Depending on the nature of the terminal conditions y can assume a series of forms $y_1, y_2, y_3, \ldots$, etc. each of which is a solution of (1) when P is given an appropriate value. Thus if the ends are "simply supported", *and if B is uniform*, a solution is obtained by giving any integral value to n in the expressions

$$y_n = \sin n\pi x/l, \quad P = P_n = n^2\pi^2 B/l^2, \tag{2}$$

where l denotes the length of the strut.

In equation (1) y stands for the deflexion away from the line of thrust. This may or may not pass through the supports, according to the nature of the terminal conditions; but in either event we may alter the definition of y to make that symbol stand for deflexion away from the line through the two ends, if in (1) we replace y by $(y+\alpha+\beta x)$ where α and β are constants. Therefore the equation

$$\frac{d^2}{dx^2}\left(B\frac{d^2y}{dx^2}\right)+P\frac{d^2y}{dx^2}=0 \tag{3}$$

will hold under the new definition of y as well as under the old: in other words, (3) is the general form of the governing equation for the strut problem.

---

* Cf. (e.g.) Southwell 1936, §§ 197–202 and Chap. XIII.

Alternatively, with the altered definition of y we may replace (1) by the equation

$$B\frac{d^2y}{dx^2}+Py=A+Cx, \tag{4}$$

which can be obtained from (3) by integration. The constants A and C will vanish when both ends are simply supported, because in that event both terms on the left-hand side must vanish for two different values of x. Then equation (1) will hold with the altered definition of y,—as was to be expected, since in these circumstances it is evident that the line of thrust must pass through both supports.

NON-DIMENSIONAL EQUATIONS, AND THE BASIC ASSUMPTION OF THE RELAXATION METHOD

5. With a view to numerical computation we now express (3) and (4) in "non-dimensional" form by writing

$$z=x/l, \quad B=B_0\beta(z), \quad \lambda=Pl^2/B_0, \quad y=Y(z), \tag{5}$$

where B_0 stands for the value of B at some representative section (e.g. at the middle or at one end). Then (4) takes the form

$$\beta(z)\,Y''+\lambda Y=\alpha+\gamma z, \tag{6}$$

and (3) becomes similarly

$$\frac{d^2}{dz^2}\{\beta(z)Y''\}+\lambda Y''=0, \tag{7}$$

dashes denoting differentiations with respect to z, and Y being a function of z whose absolute magnitude (in the nature of the problem) is indeterminate. Our problem is to determine forms for Y which, associated with particular values of λ, constitute solutions of (6) or (7) which satisfy the imposed terminal conditions. λ is the "characteristic number" (§ 1).

The assumption underlying our attack on this problem is that Y can be represented *with sufficient accuracy for practical purposes* by a finite series of independent functions Y_1, Y_2, ..., Y_n each of which satisfies severally all of the terminal conditions imposed on Y. Either end may be either (i) "simply supported", (ii) "clamped", (iii) "free", or (iv) free to move but constrained to preserve its initial direction. The mathematical statement of these four conditions is

$$\left.\begin{aligned}&\text{(i)}\ Y=Y''=0, &&\text{(ii)}\ Y=Y'=0,\\ &\text{(iii)}\ Y''=Y'''=0, &&\text{(iv)}\ Y'=\frac{d}{dz}(\beta Y'')=0.\end{aligned}\right\} \tag{8}$$

The conditions (i) are of most importance in practice, and they reduce in effect to one if (cf. § 4) we make α and γ zero in (6), thus obtaining

$$\beta(z)\,Y''+\lambda Y = 0 \tag{9}$$

as the governing equation. But in every case we shall assume that Y is representable by the series

$$Y = a_1Y_1+a_2Y_2+\ldots+a_nY_n, \tag{10}$$

in which Y_1, Y_2, …, Y_n satisfy the terminal conditions severally and are *independent* in the sense that no one can be obtained from the remainder by synthesis.

I (A). THE METHOD OF "RELAXATION BY INSPECTION"

6. In describing this method we shall postulate conditions of "simple support", so that (9) is to be solved by an expression for Y which vanishes (with Y'') at both ends. Since the same conditions are imposed on Y_1, Y_2, …, Y_n, appropriate forms are

$$Y_1 = \sin\pi z,\quad Y_2 = \sin 2\pi z,\quad \ldots,\quad Y_n = \sin n\pi z, \tag{11}$$

these being clearly "independent" in the sense of § 5.

We may say that (9) expresses the equality of a disturbing moment due to thrust, given by

$$\left.\begin{aligned}&\mu_T(\text{say}) = \lambda Y = \lambda\Sigma_n[a_n\sin n\pi z],\\ &\text{and of a restoring moment}\\ &\mu_E(\text{say}) = -\beta(z)\,Y'' = \beta(z)\,\Sigma_n[n^2\pi^2a_n\sin n\pi z]\end{aligned}\right\} \tag{12}$$

due to the stresses entailed by flexure.* Simplified by the basic assumption of § 5, our problem is to determine values of the ratios a_2/a_1, a_3/a_1, …, a_n/a_1 such that (to a sufficient degree of approximation) μ_T and μ_E balance throughout the length of the strut. *We assume that they will do so everywhere if they do so at the n sections which divide the length into* $(n+1)$ *equal parts*, and in this way we obtain n relations between the $(n-1)$ unknown ratios and the characteristic number λ whose value is required,—namely,

$$\pi^2\beta(z)\,\Sigma_n[n^2a_n\sin n\pi z] = \lambda\Sigma_n[a_n\sin n\pi z]$$

$$\text{when}\qquad z = (1, 2, \ldots, n)/(n+1). \tag{13}$$

---

[*] The double differentiation of (10) is permissible in this instance because Y'' as well as Y vanishes at either end, and will always be permissible provided that the terminal conditions are satisfied by each of Y_1, …, Y_n severally.

Only the *ratios* $a_2/a_1, \ldots$, etc., are involved, because (§ 5) the absolute magnitude of Y is indeterminate. Accordingly any one of $a_1, a_2, \ldots, a_n$ may be given the value unity.

7. Giving n the value 9 (so that the numbered sections divide the length into 10 equal parts, and at them z has the values 0·1, 0·2, …, 0·9), let us suppose that we want to determine the first (or lowest) critical value of λ. We start by assuming that $Y = Y_1$ simply (since on general grounds we know that the first mode will be a curve of one bay), and we calculate and plot Y and $-\beta(z)\,Y''$ for each of the numbered sections. Then by inspection we choose a starting value for λ such that μ_T and μ_E are approximately equal (i.e. so that (13) is approximately satisfied).

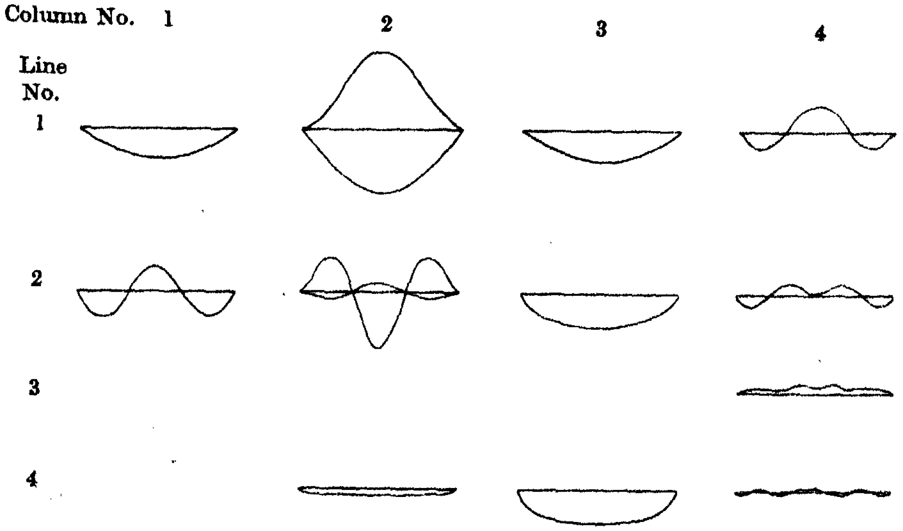


FIG. 1

Thereafter the relaxation process is best explained with the aid of diagrams. At each stage we either modify the mode by superposing (in suitable proportion) one of the "group displacements" $Y_2, Y_3, \ldots, Y_n$, or seek a closer balance between μ_T and μ_E by altering the value of λ. In fig. 1, the left-hand column shows the type of the added group displacement, the next shows the additions to μ_E and μ_T (as given by (12) for the assumed value of λ) which it entails, column 3 shows the mode (Y) and column 4 the "residual moment" ($\mu_E - \mu_T$) which result when a chosen amount of this group displacement has been added. Thus in line 1 the chosen group displacement is Y_1, and Y (third column) = Y_1 simply; values of $-\pi^2\beta(z)\,Y_1$ and of Y_1 are plotted above and below the line in the second column, and the residual moment which results when λ has been chosen (and recorded) is shown in column 4. From the form of the last diagram it is clear that Y_3 predominates in the required modification, and the second line relates to this particular group displacement: column 1 shows its nature, column 2 the corresponding distributions of $-\beta(z)\,Y_3''$ and of Y_3, column 3 and column 4 the mode and residual moment as modified by attaching a chosen value to a_3.

After several operations performed in this manner the residual moment will have a distribution of the kind shown in the third line, fourth column. Now it is evident (since $\mu_E - \mu_T$ is everywhere positive) that further reduction will be achieved most simply by increasing λY,—i.e. by increasing λ without altering the mode. Line 4 shows the result: there is no entry in column 1, and no change in column 3, but in column 2 an addition is made to λY, and the effect on the residuals is shown in column 4. Further progress will require the addition of some fairly "high harmonic" (the diagram suggests Y_7).

An example tractable by exact analysis

8. Fig. 1 is merely diagrammatic, for the reason that in fact residuals decrease too rapidly to permit the employment of a uniform scale. It is based on calculations made by one of us (K. N. E. B.) on a form of strut which had been investigated by Cowley and Levy (1918).

In that form the distribution of flexural rigidity is given by

$$\beta(z) = \{1 - 4k^2z^2\}^2, \tag{14}$$

when the origin is taken at the middle point of the strut. B_0 relating to the central section, an exact solution of the governing equation (9) for simply supported ends gives for the characteristic numbers

$$\lambda = 4k^2\left[1 + \frac{n^2\pi^2}{\left\{\log_e\left(\frac{1+k}{1-k}\right)\right\}^2}\right], \tag{15}$$

where n is integral. In particular, when $(1-k^2)^2 = 0{\cdot}2$ the first four values of λ (corresponding with $n = 1, 2, 3, 4$) are 8·153, 25·98, 55·68, 97·3. This case was selected for numerical attack.

9. Appropriate forms for $Y_1, Y_2, \ldots, Y_n$ are given in (11). Nine such functions were tabulated (for $z = 0, 0{\cdot}1, \ldots, 0{\cdot}9, 1$), together with values of $\beta(z)\,Y_n''$ for values of n from 1 to 9, as given in Tables I and II respectively. The form of each function was plotted (roughly) for use as described in § 7.

TABLE I. $Y_1, Y_2, \ldots, Y_9$ ACCORDING TO (11)

z	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9
0	0	0	0	0	0	0	0	0	0
0·1	0·3090170	0·5877853	0·8090170	0·9510565	1·00	0·9510565	0·8090170	0·5877853	0·3090170
0·2	0·5877853	0·9510565	0·9510565	0·5877853	0	−0·5877853	−0·9510565	−0·9510565	−0·5877853
0·3	0·8090170	0·9510565	0·3090170	−0·5877853	−1·00	−0·5877853	0·3090170	0·9510565	0·8090170
0·4	0·9510565	0·5877853	−0·5877853	−0·9510565	0	0·9510565	0·5877853	−0·5877853	−0·9510565
0·5	1·00	0	−1·00	0	1·00	0	−1·00	0	1·00

Y's which have finite values at the central section ($z = 0{\cdot}5$) are symmetrical about that section.
Y's which have zero values at the central section ($z = 0{\cdot}5$) are skew-symmetrical.

TABLE II. $\beta(z) \times (Y_1'', Y_2'', \ldots, Y_9'')$ ACCORDING TO (11)

z	$\beta(z)$	$n = 1$	2	3	4	5	6	7	8	9
0	0·20000	0	0	0	0	0	0	0	0	0
0·1	0·41758	−1·2736	−9·6900	−30·009	−62·715	−103·03	−141·11	−163·38	−155·04	−103·16
0·2	0·64159	−3·7220	−24·0892	−54·201	−59·552	0	133·99	295·09	385·43	301·48
0·3	0·83093	−6·6347	−31·1982	−22·808	77·126	205·02	173·53	−124·18	−499·17	−537·41
0·4	0·95626	−8·9760	−22·1900	49·927	143·616	0	−323·14	−271·83	355·04	727·06
0·5	1·00	−9·8696	0	88·826	0	−246·74	0	483·61	0	−799·44

Functions which have finite values at the central section ($z = 0{\cdot}5$) are symmetrical about that section.
Functions which have zero values at the central section ($z = 0{\cdot}5$) are skew-symmetrical.

TABLE III. RELAXATION PROCESS FOR "COWLEY AND LEVY" STRUT: FIRST CRITICAL LOAD

	Relaxation operations and multipliers	λ	$\Delta\lambda$	Points						Σ	$\lvert\Sigma\rvert$	Group displacements and multipliers	Points						Σ
				0	1	2	3	4	5				0	1	2	3	4	5	
1				0	0	0	0	0	0	0	0		0	0	0	0	0	0	0
2	$\left.\begin{matrix}-\beta(z)\,Y_1''\\-\lambda\,Y_1\end{matrix}\right\}\times 1{\cdot}00$	8·00		0	1·2736	3·7220	6·6347	8·9760	9·8696	30·4759	30·4759								
3				0	−2·4721	−4·7023	−6·4721	−7·6085	−8·0000	−29·2550		$Y_1\times 1{\cdot}00$	0	0·3090	0·5878	0·8090	0·9511	1·0000	3·6569
4				0	−1·1985	−0·9803	0·1626	1·3675	1·8696	1·2209			0	0·3090	0·5878	0·8090	0·9511	1·0000	3·6569
5	$\left.\begin{matrix}-\beta(z)\,Y_3''\\-\lambda\,Y_3\end{matrix}\right\}\times 0{\cdot}02$			0	0·6002	1·0840	0·4562	−0·9985	−1·7765	−0·6346									
6				0	−0·1294	−0·1522	−0·0494	0·0940	0·1600	−0·0770		$Y_3\times 0{\cdot}02$	0	0·0162	0·0190	0·0062	−0·0118	−0·0200	0·0096
7				0	−0·7277	−0·0485	0·5694	0·4630	0·2531	0·5093			0	0·3252	0·6068	0·8152	0·9393	0·9800	3·6665
8	$\left.\begin{matrix}-\beta(z)\,Y_5''\\-\lambda\,Y_5\end{matrix}\right\}\times 0{\cdot}002$			0	0·2061	0	−0·4100	0	0·4935	0·2896									
9				0	−0·0160	0	0·0160	0	−0·0160	−0·0160		$Y_5\times 0{\cdot}002$	0	0·0020	0	−0·0020	0	0·0020	0·0020
10				0	−0·5376	−0·0485	0·1754	0·4630	0·7306	0·7829			0	0·3272	0·6068	0·8132	0·9393	0·9820	3·6685
11	$\left.\begin{matrix}-\beta(z)\,Y_3''\\-\lambda\,Y_3\end{matrix}\right\}\times 0{\cdot}01$			0	0·3001	0·5420	0·2281	−0·4993	−0·8883	−0·3174									
12				0	−0·0647	−0·0761	−0·0247	0·0470	0·0800	−0·0385		$Y_3\times 0{\cdot}01$	0	0·0081	0·0095	0·0031	−0·0059	−0·0100	0·0048
13				0	−0·3022	0·4174	0·3788	0·0107	−0·0777	0·4270			0	0·3353	0·6163	0·8163	0·9334	0·9720	3·6733
14	$\left.\begin{matrix}-\beta(z)\,Y_5''\\-\lambda\,Y_5\end{matrix}\right\}\times 0{\cdot}001$			0	0·1030	0	−0·2050	0	0·2467	0·1447									
15				0	−0·0080	0	0·0080	0	−0·0080	−0·0080		$Y_5\times 0{\cdot}001$	0	0·0010	0	−0·0010	0	0·0010	0·0010
16				0	−0·2072	0·4174	0·1818	0·0107	0·1610	0·5637			0	0·3363	0·6163	0·8153	0·9334	0·9730	3·6743
17	$\left.\begin{matrix}-\beta(z)\,Y_7''\\-\lambda\,Y_7\end{matrix}\right\}\times 0{\cdot}001$			0	0·1634	−0·2951	0·1242	0·2718	−0·4836	−0·2193									
18				0	−0·0065	0·0076	−0·0025	−0·0047	0·0080	0·0019		$Y_7\times 0{\cdot}001$	0	0·0008	−0·0010	0·0003	0·0006	−0·0010	−0·0003
19				0	−0·0503	0·1299	0·3035	0·2778	−0·3146	0·3463			0	0·3371	0·6153	0·8156	0·9340	0·9720	3·6740
20	$\left.\begin{matrix}-\beta(z)\,Y_5''\\-\lambda\,Y_5\end{matrix}\right\}\times 0{\cdot}001$			0	0·1030	0	−0·2050	0	0·2467	0·1447									
21				0	−0·0080	0	0·0080	0	−0·0080	−0·0080		$Y_5\times 0{\cdot}001$	0	0·0010	0	−0·0010	0	0·0010	0·0010
22				0	0·0447	0·1299	0·1065	0·2778	−0·0759	0·4830	0·6348		0	0·3381	0·6153	0·8146	0·9340	0·9730	3·6750
23			0·10	0	−0·0338	−0·0615	−0·0815	−0·0934	−0·0973	−0·3675									
24		8·10		0	0·0109	0·0684	0·0250	0·1844	−0·1732	0·1155	0·4619								
25	$\left.\begin{matrix}-\beta(z)\,Y_9''\\-\lambda\,Y_9\end{matrix}\right\}\times 0{\cdot}0002$			0	0·0206	−0·0603	0·1075	−0·1454	0·1599	0·0823									
26				0	−0·0005	0·0010	−0·0013	0·0015	−0·0016	−0·0009		$Y_9\times 0{\cdot}0002$	0	0·0001	−0·0001	0·0002	−0·0002	0·0002	0·0002
27				0	0·0310	0·0091	0·1312	0·0405	−0·0149	0·1969	0·2267		0	0·3382	0·6152	0·8148	0·9338	0·9732	3·6752
28	$\left.\begin{matrix}-\beta(z)\,Y_5''\\-\lambda\,Y_5\end{matrix}\right\}\times 0{\cdot}0003$			0	0·0309	0	0·0615	0	0·0740	0·0434									
29				0	−0·0024	0	0·0024	0	−0·0024	−0·0024		$Y_5\times 0{\cdot}0003$	0	0·0003	0	−0·0003	0	0·0003	0·0003
30				0	0·0595	0·0091	0·0721	0·0405	0·0567	0·2379	0·2379		0	0·3385	0·6152	0·8145	0·9338	0·9735	3·6755
31			0·05	0	−0·0169	−0·0308	−0·0407	−0·0467	−0·0487	−0·1838									
32		8·15		0	0·0426	−0·0217	0·0314	−0·0062	0·0080	0·0541	0·1099								
33			0·01	0	−0·0034	−0·0062	−0·0081	−0·0093	−0·0097	−0·0367									
34		8·16		0	0·0392	−0·0279	0·0233	−0·0155	−0·0017	0·0174	0·1076								
35	$\left.\begin{matrix}-\beta(z)\,Y_9''\\-\lambda\,Y_9\end{matrix}\right\}\times -0{\cdot}00002$			0	−0·0021	0·0060	−0·0108	0·0145	−0·0160	−0·0084									
36				0	0·0001	−0·0001	0·0001	−0·0002	0·0002	0·0001		$Y_9\times -0{\cdot}00002$	0	—	—	—	—	—	0
37				0	0·0372	−0·0220	0·0126	−0·0012	−0·0175	0·0091	0·0905		0	0·3385	0·6152	0·8145	0·9338	0·9735	3·6755
38	$\left.\begin{matrix}-\beta(z)\,Y_7''\\-\lambda\,Y_7\end{matrix}\right\}\times -0{\cdot}00005$			0	−0·0082	0·0148	−0·0062	−0·0136	0·0242	0·0110									
39				0	0·0003	−0·0004	0·0001	0·0002	−0·0004	−0·0002		$Y_7\times -0{\cdot}00005$	0	—	—	—	—	—	0
40				0	0·0293	−0·0076	0·0065	−0·0146	0·0063	0·0199	0·0643		0	0·3385	0·6152	0·8145	0·9338	0·9735	3·6755

Starting from the assumption $Y = Y_1$ (§ 7), and with a value of λ estimated from the equation

$$\Sigma[\beta(z)\,Y''] + \lambda\Sigma[Y] = 0 \tag{16}$$

(in which both summations extend to all of the numbered points), the first three lines of the Relaxation Table III were written in. (On account of symmetry only one-half of the length had to be considered.) A summation showed (line 4) that the mode should be modified by the addition of a term a_3Y_3, and an examination on the lines of § 7 suggested a value $a_3 = 0{\cdot}02$: this gave lines 5 and 6, and a new summation (line 7) suggested further modification by means of Y_5. Five further repetitions of the procedure led to the summation in line 22, indicating the desirability of modifying λ (cf. fig. 1, line 3). An addition $\Delta\lambda = 0{\cdot}1$ was decided on, and the corresponding corrections to the residual moment (namely, $-\Delta\lambda\,.\,Y$) were inserted in line 23. (With a view to this modification account had been kept of Y on the right-hand side of Table III.)

After two more operations affecting Y, λ was again increased; and then two more operations brought the residual moments to values which were deemed negligible. The criterion employed for this purpose was the sum of the residual moments for the five points considered, taken without regard to signs (this is the quantity denoted by $|\Sigma|$ in one of the central columns). The other summation column provides an arithmetical check on the work at every stage.

10. The final value of λ in Table III is 8·16—differing by less than 0·1 % from the correct value 8·153 (§ 8). Similar treatment led to estimated values 26·1, 56·5, 100·1 for the second, third and fourth critical values—that is, to values overestimated by 0·5, 1·5, 2·9 %; and by subdividing the length into 20 instead of 10 parts the last figure was reduced to 97·5—an overestimate of 0·2 %. The starting assumption was $Y = Y_2, Y_3, Y_4$, respectively, in these three cases, and (as was to be expected) Y_2 and Y_4 called for modification only by an addition of even, Y_3 (like Y_1) only by an addition of odd harmonics. Fig. 2 shows the modes which resulted after modification.

An attempt was made to increase the accuracy of our solution for the first critical load by applying Rayleigh's theorem (§ 14) to the curve of deflexion obtained by the relaxation process. Integrations were performed by Simpson's rule and by a graphical method, but no increase in accuracy resulted.

Other types of end-condition

11. It seems safe to conclude from this example that "relaxation by inspection" will lead in all practical cases to sufficiently exact forms of Y

and values of λ; and it is an easy process to apply, since the next step is never in doubt. When other terminal conditions are imposed, appropriate changes must be made in the type solutions $Y_1, Y_2, \ldots, Y_n$. Thus if the strut is

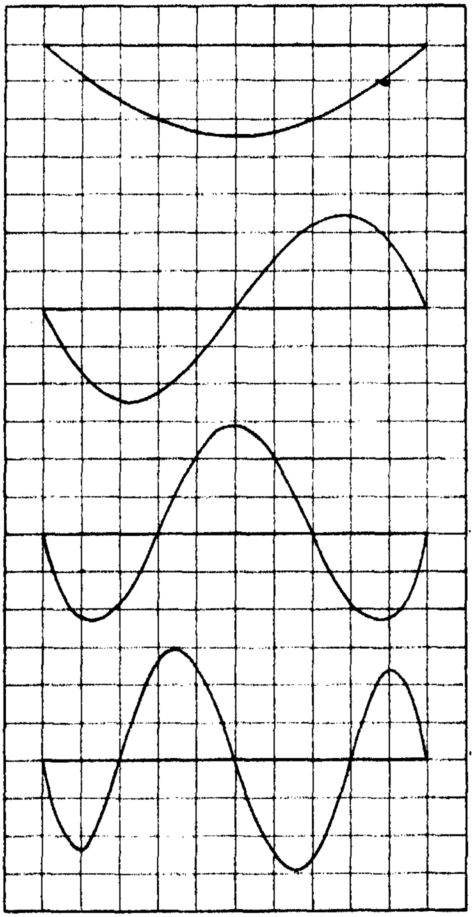


FIG. 2

"simply supported" at the end ($z=0$) and "clamped" at the end ($z=1$), appropriate forms for Y_n are given by

$$Y_n \propto \sin \alpha_n z - z \sin \alpha_n, \tag{17}$$

where α_n is the nth root of the equation

$$\sin \alpha = \alpha \cos \alpha; \tag{18}$$

when the strut is "clamped" at both ends ($z = \pm \frac{1}{2}$), appropriate forms for Y_n are given by

$$Y_n \propto \cos n\pi - \cos 2n\pi z \quad (n \text{ being integral}) \tag{19}$$

and by
$$\left.\begin{aligned} &Y_n \propto 2z \sin \tfrac{1}{2}\alpha_n - \sin \alpha_n z, \\ &\text{where } \alpha_n \text{ is the } n\text{th root of the equation} \\ &2 \sin \tfrac{1}{2}\alpha = \alpha \cos \tfrac{1}{2}\alpha. \end{aligned}\right\} \tag{20}$$

A solution of (20) comes between every solution of (19). All of these results are exact solutions for a *uniform* strut.

12. Contrary to expectation, less satisfactory results were obtained when the method of "relaxation by inspection" was applied (by K. N. E. B.) to the problem of transverse vibrations in a non-uniform rod. The natural frequencies depend both on mass and flexural rigidity, neither of which (in general) is independent of z: on this account "liquidation" proceeded more slowly, and it was often difficult to decide the nature of the next step. Accordingly a systematic procedure was developed, based on Rayleigh's theorem regarding the stationary property of the normal modes. Although hardly needed in relation to struts, the procedure is most simply explained in that connexion, and will be so presented here.

I (B). THE STRUT PROBLEM TREATED BY "SYSTEMATIC RELAXATION"

13. When both terms of (3), multipled by $\frac{1}{2}y$, are integrated over the range $0 \leqslant x \leqslant l$, the resulting equation is

$$\frac{1}{2}\int_0^l B y''^2 dx = -\frac{1}{2} P \int_0^l y y'' dx = \frac{1}{2} P \int_0^l y'^2 dx, \tag{21}$$

provided that at each end the terminal conditions are one or other of those given in (8) of § 5—i.e. do not permit either an input or output of energy.* Equation (21) states that the strain-energy of flexure equals the work done by end-thrust on account of the curvature (whereby the two ends of the strut approach one another):† it could thus have been deduced from energy principles, and it will be satisfied not only when y and P constitute an exact solution of (3) but also when the mode is imposed by constraints, provided that *on the whole* these do no work. On that understanding the thrust is given by

---

* At a "free" end (Case (iii) of § 5) y as defined in § 4 will vanish because this end is on the line of thrust. For stability the other end must be clamped, so that there y' is zero.

† Cf. Southwell 1936, §§ 480–1.

$$P = \int_0^l By''^2\,dx \bigg/ \int_0^l y'^2\,dx, \tag{22}$$

dashes now denoting differentiations with respect to x.

When y is identified with y_n as defined in § 4, equation (22) reduces to

$$P = P_n,$$

as was to be expected. Rayleigh's theorem asserts further that *when y is identified with y_n, P as given by* (22) *is stationary for all variations of y.*

To prove the theorem we write

$$N = \int_0^l By''^2\,dx, \quad D = \int_0^l y'^2\,dx \tag{23}$$

in (22). Then we have

$$\delta P = \delta(N/D) = \frac{1}{D}\left(\delta N - \frac{N}{D}\delta D\right) = \frac{1}{D}(\delta N - P\,\delta D), \tag{24}$$

showing that δP will vanish if for all permitted variations

$$2\int_0^l By''\,\delta y''\,dx = \delta N = P\,\delta D = 2P\int_0^l y'\,\delta y'\,dx.$$

Integrating by parts and taking account of the terminal conditions, we deduce that P will be stationary if

$$\int_0^l \delta y \frac{d^2}{dx^2}(By'' - Py)\,dx = 0 \tag{25}$$

for all permitted variations δy; and evidently this will be the fact if y and P are identified with y_n and P_n, since in that event (3) is satisfied.

14. The argument can be taken further. For the ratio on the right of (22) is necessarily positive, and accordingly its least stationary value—i.e. P_1, the lowest "critical load"—must be a *minimum*. Thus according to Rayleigh's theorem, if y is a close approximation to y_1, then (22) will give a very close approximation to P_1, and *the error will be one of excess*.

Rayleigh's theorem is the basis of the approximate methods now to be described. We want to determine the form of y_1 and the value of P_1, and we assume for this purpose (cf. § 5) that y_1 can be represented *with sufficient accuracy* by a synthesis of type-solutions Y_1, Y_2, ..., etc., *finite in number*. In the notation of § 5 we may replace (22) and (23) by

$$\lambda = N/D, \quad N = \int_0^1 \beta(z)\, Y''^2\, dz, \quad D = \int_0^1 Y'^2\, dz, \tag{26}$$

dashes now denoting differentiations with respect to z; and our problem is to find values of the ratios a_2/a_1, a_3/a_1, ..., a_n/a_1 such that λ as given by (26) will have the lowest value which is possible when Y is replaced by the series

$$Y = a_1 Y_1 + a_2 Y_2 + \ldots + a_n Y_n. \qquad (10)\,\textit{bis}$$

Determination of the first critical load

15. Without loss of generality we may give to a_1 the value unity, since we know (§ 5) that the absolute magnitude of Y is indeterminate. Then (consistently with our assumption) we have $(n-1)$ relations to be satisfied, namely,

$$\frac{\partial \lambda}{\partial a_r} = 0 \quad (r = 2, 3, \ldots, n), \tag{27}$$

when λ is given by (26) and Y by (10) with $a_1 = 1$.

Evidently

$$\left.\begin{aligned}
&\frac{\partial N}{\partial a_r} = 2\int_0^1 \beta(z)\, Y'' \frac{\partial Y''}{\partial a_r}\, dz = 2\int_0^1 \beta(z)\, Y'' Y_r''\, dz,\\
&\frac{\partial D}{\partial a_r} = 2\int_0^1 Y' \frac{\partial Y'}{\partial a_r}\, dz = 2\int_0^1 Y' Y_r'\, dz = -2\int_0^1 Y'' Y_r\, dz,\\
&\text{in virtue of the terminal conditions; and in the same way}\\
&\frac{\partial^2 N}{\partial a_r^2} = 2\int_0^1 \beta(z)\, Y_r''^2\, dz, \quad \frac{\partial^2 N}{\partial a_r \partial a_s} = 2\int_0^1 \beta(z)\, Y_r'' Y_s''\, dz,\\
&\frac{\partial^2 D}{\partial a_r^2} = -2\int_0^1 Y_r'' Y_r\, dz = 2\int_0^1 Y_r'^2\, dx, \quad \frac{\partial^2 D}{\partial a_r \partial a_s} = 2\int_0^1 Y_r' Y_s'\, dz.
\end{aligned}\right\} \tag{28}$$

So ($Y_1, Y_2, \ldots, Y_n$ being definite) the first derivatives of N and D are linear in $a_2, a_3, \ldots, a_n$, and all of the second derivatives have constant values which can be determined once for all. Moreover, if

$$\left.\begin{aligned} & & -2A_r \equiv D\frac{\partial\lambda}{\partial a_r} &= \frac{\partial N}{\partial a_r} - \lambda\frac{\partial D}{\partial a_r}, \\ &\text{then} & -2\frac{\partial A_r}{\partial a_r} &= \frac{\partial^2 N}{\partial a_r^2} - \lambda\frac{\partial^2 D}{\partial a_r^2} - \frac{\partial\lambda}{\partial a_r}\frac{\partial D}{\partial a_r} \\ & & &= \frac{\partial^2 N}{\partial a_r^2} - \lambda\frac{\partial^2 D}{\partial a_r^2} + 2\frac{A_r}{D}\frac{\partial D}{\partial a_r}, \\ &\text{and similarly} & -2\frac{\partial A_r}{\partial a_s} &= \frac{\partial^2 N}{\partial a_r\partial a_s} - \lambda\frac{\partial^2 D}{\partial a_r\partial a_s} + 2\frac{A_s}{D}\frac{\partial D}{\partial a_r}. \end{aligned}\right\} \qquad (29)$$

We want to find values of $a_2, a_3, \ldots, a_n$ which will make all of $A_2, \ldots, A_n$ zero. Therefore in the terminology of the relaxation method we may regard the a's as "displacements", the A's as "residual forces", and derivatives like $\partial A_r/\partial a_r$, $\partial A_r/\partial a_s$ as "influence coefficients". Compared with problems treated hitherto the problem treated here is novel in that *the influence coefficients have values which are not constant*, since their expressions involve the A's and also λ.*

16. On the other hand we know from Rayleigh's theorem (§ 14) that starting with a reasonably close approximation to the correct mode we shall start with a very close approximation to λ, and thereafter (since we are working in the neighbourhood of its minimum value) λ will alter very slowly. The A's will be small for the same reason, also the a's and quantities of the type $A_r/D.\partial D/\partial a_r$ in (29). Accordingly the "influence coefficients" will not alter widely as the work proceeds, and on this account their values will require correction only at occasional stages in the relaxation process. Now it is a feature of the relaxation method that in effect it starts afresh at every operation, treating previous results as a "trial solution" which calls for modification: therefore we shall lose no accuracy in the final stages because in the early stages we have used influence coefficients which are not quite exact.

---

* Cf. Black and Southwell 1938, § 4. If $\frac{1}{2}\lambda$ were taken as the analogue of "total potential energy" $(U+V)$, the residual forces would be given by

$$-2A_r = \partial\lambda/\partial a_r,$$

etc., and Maxwell's reciprocal relations would hold in relation to the influence coefficients (of type $\partial^2\lambda/\partial a_r\partial a_s$). By inserting the multiplier D in (29) we simplify the resulting formulae with a loss of the reciprocal relations, which are now satisfied only approximately, because D is only approximately constant.

At the close of any stage, when the "forces" A_r, A_s have been brought practically to zero, they will make no appreciable contribution to the expressions (29) for $\partial A_r/\partial a_r$, $\partial A_r/\partial a_s$. If then we neglect them throughout the stage, our results will provide good starting assumptions for the next stage, and we shall be using constant influence coefficients such that

$$\frac{\partial A_r}{\partial a_r} < 0, \quad \frac{\partial A_r}{\partial a_s} = \frac{\partial A_s}{\partial a_r}. \tag{30}$$

The "framework analogy" will be exact, and arguing on the lines of Part II (Black and Southwell 1938) it is easy to show that our results will converge.

17. We have still to define the forms of the type-solutions $Y_1, Y_2, \ldots, Y_n$. In the strut problem we may postulate that they are the 1st, 2nd, ..., nth modes which satisfy both the governing equation (7) and the imposed terminal conditions of our problem when the flexural rigidity is uniform. These are easily calculable,* and under any terminal conditions which do not permit a transfer of energy (§ 13) it is easy to show (Southwell 1936, § 485) that they satisfy "conjugate relations" expressed by

$$\int_0^1 Y_r'' Y_s'' \, dz = \int_0^1 Y_r' Y_s' \, dz = 0 \quad (r \neq s). \tag{31}$$

This means, according to (28), that all derivatives of the type $\partial^2 D/\partial a_r \partial a_s$ will vanish and (provided that $\beta(z)$, as will usually be the fact, does not vary widely) that derivatives of the type $\partial^2 N/\partial a_r \partial a_s$ will be small in comparison with $\partial^2 N/\partial a_r^2$, etc. *Moreover (for any practical form of strut)* Y_1 *will be a fairly close approximation to the required first mode* y_1: therefore in the resulting "solution", and so at every stage in a systematic process of relaxation, $a_2, a_3, \ldots, a_N$ will have values which are small in relation to unity.

18. On this understanding in regard to $Y_1, Y_2, \ldots, Y_n$ we may proceed to calculate, once for all, the quantities

$$I_{rr} = \int_0^1 Y_r'^2 \, dz, \quad J_{rs} = \int_0^1 \beta(z) \, Y_r'' Y_s'' \, dz, \quad J_{rr} = \int_0^1 \beta(z) \, Y_r''^2 \, dz, \tag{32}$$

of which the I's are standard for all struts. Then by (10) and (26) we have

$$\left. \begin{aligned} N &= a_1^2 J_{11} + a_2^2 J_{22} + \ldots + a_n^2 J_{nn} + 2(a_1 a_2 J_{12} + \ldots + a_r a_s J_{rs} + \ldots), \\ D &= a_1^2 I_{11} + a_2^2 I_{22} + \ldots + a_n^2 I_{nn}, \end{aligned} \right\} \tag{33}$$

in which a_1 (§ 15) is to be given the value unity.

---

* Thus when both ends are simply supported the Y's will be circular functions of z (cf. § 4). Their forms under other terminal conditions have been given in (17)–(20) of § 11.

In seeking the first critical load we began by assuming that $Y = Y_1$ simply, so that all the other a's vanish. Then we have initially

$$\left.\begin{aligned} &N = J_{11}, \quad D = I_{11}, \quad \lambda = N/D = J_{11}/I_{11}, \\ \text{and} \qquad &\frac{\partial N}{\partial a_r} = 2J_{1r}, \quad \frac{\partial D}{\partial a_r} = 0. \end{aligned}\right\} \tag{34}$$

Also we have

$$\left.\begin{aligned} &\frac{\partial^2 N}{\partial a_r^2} = 2J_{rr}, \quad \frac{\partial^2 N}{\partial a_r \partial a_s} = 2J_{rs}, \quad \frac{\partial^2 D}{\partial a_r^2} = I_{rr}, \quad \frac{\partial^2 D}{\partial a_r \partial a_s} = 0, \\ \text{so that } \textit{initially} \qquad &-\frac{\partial A_r}{\partial a_r} = J_{rr} - \lambda I_{rr}, \quad -\frac{\partial A_r}{\partial a_s} = J_{rs}, \end{aligned}\right\} \tag{35}$$

λ having its value according to (34).

We insert these initial values in an Operations Table, and we start a Relaxation Table by inserting, in the first line, values of the quantities typified by

$$A_r = -\frac{1}{2}\left(\frac{\partial N}{\partial a_r} - \lambda \frac{\partial D}{\partial a_r}\right) = -J_{1r} \text{ according to (34).} \tag{36}$$

The relaxation process then proceeds as before, and when the "residual forces" are deemed sufficiently small we have interim values of $a_2, \ldots, a_n$ which can be used to calculate revised values of λ, of the residual forces, and of the "influence coefficients" wherewith we can start the second stage. The formulae (33) give the new approximation to λ ($= N/D$); the residual forces (of type A_r) come from (29) with

$$\left.\begin{aligned} &\frac{1}{2}\frac{\partial N}{\partial a_r} = J_{1r} + a_2 J_{2r} + a_3 J_{3r} + \ldots + a_n J_{nr}, \\ &\frac{1}{2}\frac{\partial D}{\partial a_r} = a_r I_{rr}, \end{aligned}\right\} \tag{37}$$

by (28), (10) and (32); and the new values of the influence coefficients are given by the equations

$$\left.\begin{aligned} -\frac{\partial A_r}{\partial a_r} &= J_{rr} - \frac{a_r I_{rr}}{D}\frac{\partial N}{\partial a_r} - \lambda I_{rr}\left(1 - 2\frac{a_r^2}{D} I_{rr}\right), \\ -\frac{\partial A_r}{\partial a_s} &= J_{rs} - \frac{a_r I_{rr}}{D}\frac{\partial N}{\partial a_s} + 2\lambda \frac{a_r a_s}{D} I_{rr} I_{ss}, \end{aligned}\right\} \tag{38}$$

according to (29), (35) and (37). Relaxation is effected as before, and still closer approximation can be sought, if desired, by again revising the values of λ, etc. and proceeding to a further stage.

Determination of the higher critical loads

19. Provided that Y_r is a close approximation to the correct rth mode, the same procedure (with a_r instead of a_1 (§ 15) kept unity) will yield a close estimate of P_r. But in general the effect of non-uniform flexural rigidity will be to make the approximation of Y_1, Y_2, ..., etc., successively worse, so that the a's which have to be varied in the relaxation process are no longer small; and when two of the a's are comparable it is clearly possible that our suggested procedure may not lead to the mode and characteristic number that are sought. We have shown (§ 13) that every critical loading P_n has the stationary property; but only P_1 is a *minimum* of the ratio (22), and the advantage resulting from this circumstance will be lost when our methods (as presented above) are applied to the determination of "higher modes".

But now let y_r, y_s, as in § 4, stand for exact solutions of (3) which are associated with different values of P. Then it is easy to establish the conjugate relations

$$\int_0^l By_r''y_s''\,dx = \int_0^l y_r'y_s'\,dx = 0 \tag{39}$$

(Southwell 1936, § 485); and it follows that a given mode y, regarded as composed of exact solutions y_1, y_2, ..., etc., combined in suitable proportions, will have no component y_1 if

$$\int_0^l y'y_1'\,dx = 0. \tag{40}$$

Suppose that y is restricted by this condition. Then it can be shown as in § 13 that P according to (22) will have a stationary value when y is identified with any exact solution of (3); but now, since y cannot have the form y_1, the lowest stationary value of P will be found when y is identified with y_2. Accordingly (by the argument of § 14) the minimum value of P is P_2, and if y is a reasonably close approximation to y_2 then (22) will give a very close approximation to P_2 and *the error will be one of excess*. Using this result, we can still take advantage (when necessary) of Rayleigh's *minimal* theorem.

20. In our "non-dimensional" notation the relation (40) is equivalent to

$$\int_0^1 Y\overline{Y}_1''\,dz = 0, \tag{41}$$

in virtue of the terminal conditions, $\overline{Y}_1$ denoting the non-dimensional form of y_1: if (41) is satisfied, then Y will have no component $\overline{Y}_1$. Thus *when the form of $\overline{Y}_1$ is known* it is a simple matter to arrange that an assumed form Y shall have this feature: we have only to replace (10) by the expression

$$Y = a_2(Y_2-\alpha_2Y_1)+a_3(Y_3-\alpha_3Y_1)+\ldots+a_n(Y_n-\alpha_nY_1), \tag{42}$$

which is equivalent to (10) if $a_1+a_2\alpha_2+\ldots+a_n\alpha_n=0$, and to arrange that every expression of the type $(Y_r-\alpha_r Y_1)$ shall satisfy (41) severally. This will be the case if

$$\alpha_r\int_0^1 Y_1\,\overline{Y}_1''\,dz=\int_0^1 Y_r\,\overline{Y}_1''\,dz. \tag{43}$$

When the α's have been thus determined, we have in (42) an expression automatically satisfying the imposed condition *at the cost of one arbitrary coefficient*; for since (§ 5) the magnitude of Y is indeterminate, no generality is lost if a_2 is given the value 1.

21. Proceeding on these lines we shall take advantage of the circumstance that by relaxation as described in §§ 15–18 we determine with close accuracy not only λ_1 but also the associated mode $\overline{Y}_1$. Applied to (42) relaxation methods will determine $\overline{Y}_2$ with an accuracy depending almost entirely on the number of the arbitrary coefficients $a_3, \ldots, a_n$: then by the same procedure, knowing $\overline{Y}_1$ and $\overline{Y}_2$ we can proceed to determine $\overline{Y}_3$, and so on. It is easy to add new type-solutions to our series if specially high accuracy is desired; but in practice we can usually be content with less accuracy in respect of higher modes.

For struts which are symmetrical the modes will fall into two distinct classes, involving respectively forms $Y_1, Y_3, \ldots$, etc. which are symmetrical, and forms $Y_2, Y_4, \ldots$, etc. which are skew-symmetrical with respect to the centre of the strut. Then knowing λ_1 and $\overline{Y}_1$ we are in a position to determine λ_3 and $\overline{Y}_3$; knowing λ_2 and $\overline{Y}_2$ we are in a position to determine λ_4 and $\overline{Y}_4$; and so on.

An example tractable by exact analysis

22. In illustration of these methods we shall consider Cowley and Levy's strut (§ 8) in which

$$\beta(z)=\{1-4k^2z^2\}^2, \tag*{(14) bis}$$

where

$$(1-k^2)^2=0{\cdot}2,$$

and the ends are simply supported. Appropriate modes are (cf. § 17 and footnote)

$$Y_1=\sin\pi z,\quad Y_2=\sin 2\pi z,\quad \ldots,\quad Y_n=\sin n\pi z, \tag*{(11) bis}$$

and for these forms the I's and J's according to (32) are recorded (for "odd" modes) in Table IV. On account of the form of $\beta(z)$ the integrations presented no difficulty: the I's are standard values applicable to all struts, I_{rr} having the expression $\frac{1}{2}r^2\pi^2$.

TABLE IV. I AND J INTEGRALS FOR "COWLEY AND LEVY" STRUT

	I_{rr}						
$r=$	1	3	5	7	9	11	13
$I_{rr}=$	4·9348	44·413	123·37	241·81	399·72	597·11	833·98

	J_{rr} and J_{rs}						
$s=$ \ $r=$	1	3	5	7	9	11	13
1	42·281	−63·875	−21·540	−13·1388	−9·5777	−7·5836	−6·2972
3	−63·875	2780·04	−1851·08	−550·36	−327·36	−236·39	−188·46
5	−21·540	−1851·08	21209·4	−10299·7	−2729·55	−1543·62	−1092·1
7	−13·1388	−550·36	−10299·7	81238	−33614	−8230·8	−4457·4
9	−9·5777	−327·36	−2729·55	−33614	221728	−83280·$_5$	−19280·3
11	−7·5836	−236·39	−1543·62	−8230·8	−83280·$_5$	494492	−174067
13	−6·2972	−188·46	−1092·1	−4457·4	−19280·3	−174067	964302

This being a symmetrical example as discussed in § 21, it will suffice to describe the investigation (by D. G. C.) of "odd" modes leading to estimates of $\lambda_1, \lambda_3, \ldots$, etc. For the determination of λ_1, starting from the assumption $Y = Y_1$ (§ 18), we had initial forces $\bar{A}_3, \bar{A}_5, \ldots, \bar{A}_9$ given by $J_{13}, J_{15}, \ldots, J_{19}$ of Table IV, and an initial value $J_{11}/I_{11} = 8{\cdot}5679$ for λ. Having calculated influence coefficients from the approximate expressions (35), we "relaxed" until A_3, A_5 had been reduced from initial magnitudes of 63·875 and 21·540 to magnitudes less than 0·01. It was not worth while, in this preliminary stage, to relax A_7 and A_9 or to permit displacements a_7, a_9: the other "displacements" were

$$a_3 = 0{\cdot}02953_5, \quad a_5 = 0{\cdot}00378,$$

and for these values (with a_7 and a_9 taken as zero) new values for λ, for the residual forces, and for the influence coefficients were calculated from (33), (37) and (38) with the terms involving $1/D$ suppressed (cf. § 16). The results were

$$\lambda = 8{\cdot}1720,$$

$$A_3 = -0{\cdot}521, \quad A_5 = -0{\cdot}186, \quad A_7 = 68{\cdot}348, \quad A_9 = 29{\cdot}570,$$

with influence coefficients as given in Table V.

TABLE V. VALUES OF $\partial A_r/\partial a_s$ (APPROXIMATE)

$s=$ \ $r=$	3	5	7	9
3	−2417·10	1851·08	550·36	327·36
5	1851·08	−20201·2	10299·7	2729·55
7	550·36	10299·7	−79262	33614
9	327·36	2729·55	33614	−218461

Relaxation was then continued for a second "stage". At its termination the final "forces" were

$$A_3 = -0{\cdot}003, \quad A_5 = 0{\cdot}000, \quad A_7 = -0{\cdot}002, \quad A_9 = 0{\cdot}002,$$

and the final "displacements" were

$$a_3 = 0{\cdot}03008, \quad a_5 = 0{\cdot}00441, \quad a_7 = 0{\cdot}001079_5, \quad a_9 = 0{\cdot}000310.$$

Then using (33) again we deduced the value 8·1553 for λ, and found as residual forces (to be used in starting a third "stage")

$$A_3 = -0{\cdot}008, \quad A_5 = -0{\cdot}004, \quad A_7 = -0{\cdot}019, \quad A_9 = 0{\cdot}015.$$

Since it was evident that further relaxation would not affect any digit in the figures for a_3, a_5, a_7, a_9, closer approximation was not attempted. The estimate 8·1553 for λ, compared with the correct value 8·153 (§ 8), is less than 0·03 % in error, and the estimate of the associated mode, as shown in the first section of Table VI, agrees very closely with the exact solution.

TABLE VI. MODES DETERMINED BY SYSTEMATIC RELAXATION

	First mode		Third mode		Fifth mode	
$z \times 18$	By relaxation method	By exact analysis	By relaxation method	By exact analysis	By relaxation method	By exact analysis
0	0	0	0	0	0	0
1	0·1986	0·1992	0·5313	0·5415	0·6688	0·7311
2	0·3832	0·3832	0·8155	0·8155	0·5702	0·5338
3	0·5460	0·5462	0·7794	0·7660	−0·1737	−0·2212
4	0·6844	0·6846	0·5246	0·5062	−0·8263	−0·8165
5	0·7982	0·7982	0·1154	0·1108	−0·9097	−0·8936
6	0·8866	0·8866	−0·3061	−0·3103	−0·4856	−0·4676
7	0·9495	0·9496	−0·6741	−0·6727	0·1664	0·1999
8	0·9873	0·9874	−0·9170	−0·9151	0·7570	0·7756
9	1	1	−1	−1	1	1

To examine the importance of including several terms in the series (10), relaxation was also effected with neglect of A_9 and a_9. The resulting estimate of λ was 8·164,—i.e. the error was increased to above 1 %: this result suggests that very high accuracy could be obtained, if desired, by increasing the number of the variable parameters.

23. It is not necessary to describe in detail the exactly similar work which led to estimates of λ_3, λ_5, λ_7, λ_9. Proceeding on the lines of §§ 19–20 (with a loss of one arbitrary parameter at every advance to a "higher mode"), and

ending every series of type (42) with the term involving Y_9, we obtained for λ the values shown in column 3 of Table VII: neglecting A_9 and Y_9 as described above we obtained the values given in column 5.

TABLE VII. CRITICAL LOADS DETERMINED BY SYSTEMATIC RELAXATION

1	2	3	4	5	6	7	8
Order of critical load	Exact solution	By relaxation method with 5 parameters	Error %	By relaxation method with 4 parameters	Error %	With lower modes eliminated (§ 20) and 5 parameters	Error %
λ_1	8·153	8·155	0·025	8·164	0·14	8·153	0·025
λ_3	55·68	55·89	0·4	56·35	1·2	55·85	0·3
λ_5	150·8	155·3	3	160·6	6·5	154·0	2·1
λ_7	293·4	315·8	8	350·4	20	—	—
λ_9	484	595	24	—	—	—	—

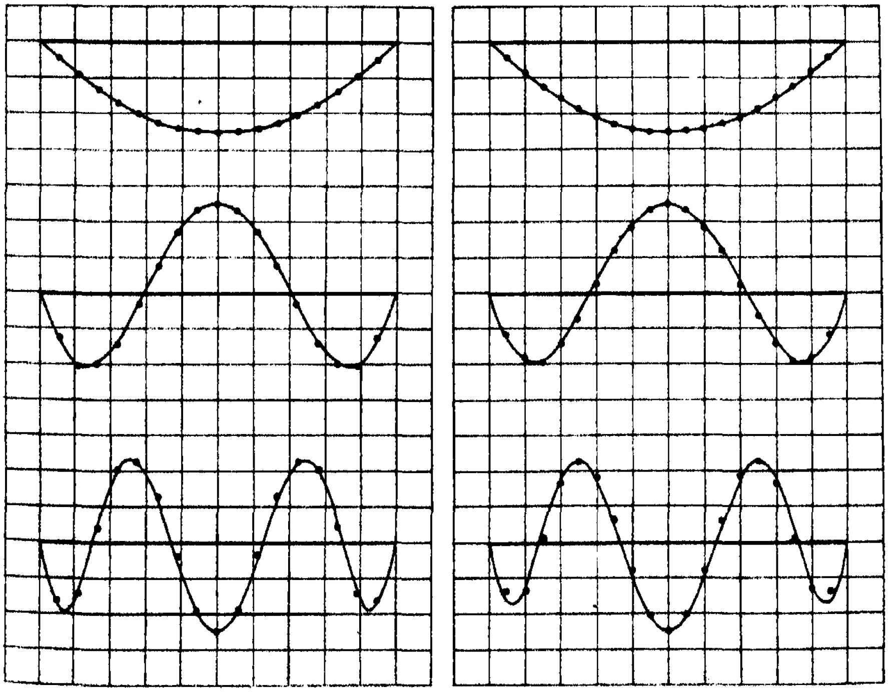


FIG. 3 FIG. 4

First, Third and Fifth Normal Modes for Strut (Fig. 3) and Vibrating Rod (Fig. 4). The curves are exact solutions: points show values obtained by the Relaxation Process.

The loss of accuracy entailed by a decrease in the number of the arbitrary parameters affords further support for the suggestion made above (§ 22).

It led us to recalculate λ_3 and λ_5 with series extending to Y_{11} and Y_{13} respectively (i.e. with the number of arbitrary parameters kept constant), and as expected the accuracy was considerably improved. Columns 7 and 8 of Table VII relate to the resulting estimates.

Dispensing with our device for eliminating (approximately) lower modes from the forms investigated, we found that in this instance the accuracy was thereby improved: with the original five a's available, λ_3 and λ_5 were estimated with errors of 0·35 and 2·05 % respectively.*

The conclusion seems legitimate that Relaxation Methods provide a means of determining critical loads, *together with the associated modes of distortion,* to any degree of accuracy which is likely to be needed in practical work. The figures in Table VI give an impression of sensible error in the determination of the third and fifth modes; but fig. 3 shows that the discrepancies are in fact small.

PART II. "WHIRLING SPEEDS", OR NATURAL FREQUENCIES OF TRANSVERSE VIBRATION, FOR A ROD OF NON-UNIFORM SECTION

THE GOVERNING EQUATION

24. When free transverse vibrations are executed by an initially straight rod, the governing equation can be shown to be†

$$\frac{d^2}{dx^2}\left(B\frac{d^2y}{dx^2}\right) = mp^2y, \tag{44}$$

where B and x have the same significance as in § 4, m stands for the mass per unit length, and $y \cos pt$ is the instantaneous deflexion, away from the position of equilibrium, of a section defined by x. ($T = 2\pi/p$ is the period of a complete vibration.)

Depending on the nature of the terminal conditions y can assume a series of forms $y_1, y_2, \ldots$, etc. each of which is a solution of (44) when p^2 is given an appropriate value. Thus if the ends are "simply supported", *and if B and m are uniform*, solutions are obtained by giving integral values to n in the expressions

$$y_n = \sin n\pi x/l, \quad p^2 = p_n^2 = n^4\pi^4 B/ml^4, \tag{45}$$

l denoting the length of the rod between supports.

---

* It was implied in § 19 that this device will not always be needed. The question evidently turns on the accuracy with which Y_3, Y_5, ... etc. represent the higher normal modes.

† Cf. (e.g.) Southwell 1936, Chap. XIV. Replacing p by ω, we have the governing equation for a shaft which "whirls" with an angular velocity ω.

RAYLEIGH'S THEOREM REGARDING THE NATURAL FREQUENCIES

25. If both sides of (44), multiplied by $\frac{1}{2}y$, are integrated "by parts" over the range $0 \leqslant x \leqslant l$, then whatever be the nature of the terminal conditions, provided only that they do not permit an input or output of energy, the equation

$$\frac{1}{2}\int_0^l B\left(\frac{d^2y}{dx^2}\right)^2 dx = \frac{1}{2}p^2\int_0^l my^2\,dx \tag{46}$$

results. This may be interpreted as stating that the maximum values of the total strain and kinetic energies are equal in a free vibration: it could have been deduced from energy principles, and will be satisfied not only when y and p^2 constitute an exact solution of (44), but also when the mode is imposed by constraints, provided that *on the whole* these do no work. On that understanding the frequency will be given by (46) in the form

$$p^2 = \int_0^l By''^2\,dx \Big/ \int_0^l my^2\,dx, \tag{47}$$

dashes denoting differentiations with respect to x.

When y is identified with y_n as defined in (45) of § 24, equation (47) reduces to

$$p^2 = p_n^2,$$

as was to be expected. Rayleigh's theorem asserts further that *when y is identified with y_n, p^2 as given by* (47) *is stationary for all variations of y.*

To prove the theorem we may write

$$N = \int_0^l By''^2\,dx, \quad D = \int_0^l my^2\,dx \tag{48}$$

in (47). Then an argument on the lines of § 13 shows that p^2 will be stationary

if
$$\int_0^l \delta y\left\{\frac{d^2}{dx^2}(By'') - mp^2y\right\}dx = 0,$$

for all permitted variations δy; and this condition will be satisfied when y and p^2 are identified with y_n and p_n^2, since in that event (44) is satisfied.

26. We can take the argument further in the manner of § 14, showing that if y is a close approximation to y_1, then (47) gives a very close approximation to p_1^2 *and the error is one of excess*; and on this result we can base an

approximate treatment on the lines of §§ 15–21. The "non-dimensional" relation corresponding with (47) is

$$\lambda = N/D, \quad N = \int_0^1 \beta(z)\, Y''^2\, dz, \quad D = \int_0^1 \mu(z)\, Y^2\, dz, \tag{49}$$

where

$$z = x/l, \quad y = Y(z), \quad B = B_0\beta(z), \quad m = m_0\mu(z), \quad \lambda = m_0 p^2 l^4/B_0, \tag{50}$$

B_0, m_0 standing for the values of B and m at some representative section. If as before we assume that Y can be closely represented by the series (10), then the problem of approximating to y_1 reduces to the finding of values for $a_2, a_3, \ldots, a_n$ such that λ has its minimum value. We have to satisfy the relations

$$\frac{\partial\lambda}{\partial a_r} = 0 \quad (r = 2, 3, \ldots, n), \tag*{(27) bis}$$

λ being given by (49) and Y by (10) with $a_1 = 1$.

27. The subsequent treatment follows, *mutatis mutandis*, that of §§ 15–21. We have according to (49)

$$\left.\begin{aligned} \frac{\partial N}{\partial a_r} &= 2\int_0^1 \beta(z)\, Y'' Y_r''\, dz, & \frac{\partial D}{\partial a_r} &= 2\int_0^1 \mu(z)\, Y Y_r\, dz, \\ \frac{\partial^2 N}{\partial a_r^2} &= 2\int_0^1 \beta(z)\, Y_r''^2\, dz, & \frac{\partial^2 N}{\partial a_r \partial a_s} &= 2\int_0^1 \beta(z)\, Y_r'' Y_s''\, dz, \\ \frac{\partial^2 D}{\partial a_r^2} &= 2\int_0^1 \mu(z)\, Y_r^2\, dz, & \frac{\partial^2 D}{\partial a_r \partial a_s} &= 2\int_0^1 \mu(z) Y_r Y_s\, dz, \end{aligned}\right\} \tag{51}$$

and we may insert these expressions in the formulae (29) for "residual forces" and for "influence coefficients". As before our problem is to find values of $a_2, a_3, \ldots, a_n$ which make all of $A_2, A_3, \ldots, A_n$ zero. The type-solutions $Y_1, Y_2, \ldots, Y_n$ (which we have not yet defined in relation to the vibration problem) may conveniently be given the same forms as before (§ 17), and having chosen them we may proceed to calculate, once for all, the quantities

$$\left.\begin{aligned} I_{rs} &= \int_0^1 \mu(z) Y_r Y_s\, dz \quad (\text{including } I_{rr}), \\ J_{rs} &= \int_0^1 \beta(z)\, Y_r'' Y_s''\, dz \quad (\text{including } J_{rr}). \end{aligned}\right\} \tag{52}$$

(Since its integrand contains $\mu(z)$ as a factor, it is no longer possible to arrange that $I_{rs} = 0$ when $r \neq s$.)

We have according to (10) and (49)

$$\left.\begin{aligned} N &= a_1^2 J_{11} + a_2^2 J_{22} + \dots + a_n^2 J_{nn} + 2(a_1 a_2 J_{12} + \dots + a_r a_s J_{rs} + \dots), \\ D &= a_1^2 I_{11} + a_2^2 I_{22} + \dots + a_n^2 I_{nn} + 2(a_1 a_2 I_{12} + \dots + a_r a_s I_{rs} + \dots). \end{aligned}\right\} \quad (53)$$

In seeking an estimate of λ_1 we may start as before by assuming that $Y = Y_1$ simply, so that all of $a_2, a_3, \dots, a_n$ vanish. Then we have initially

$$\left.\begin{aligned} & N = J_{11}, \quad D = I_{11}, \quad \lambda = N/D = J_{11}/I_{11}, \\ \text{and} \quad & \frac{\partial N}{\partial a_r} = 2J_{1r}, \quad \frac{\partial D}{\partial a_r} = 2I_{1r}, \\ \text{whence} \quad & -A_r = J_{1r} - \lambda I_{1r}, \quad \text{by (29).} \end{aligned}\right\} \quad (54)$$

Also we have *always*

$$\left.\begin{aligned} & \frac{\partial^2 N}{\partial a_r^2} = 2J_{rr}, \quad \frac{\partial^2 N}{\partial a_r \partial a_s} = 2J_{rs}, \quad \frac{\partial^2 D}{\partial a_r^2} = I_{rr}, \quad \frac{\partial^2 D}{\partial a_r \partial a_s} = I_{rs}, \\ \text{so that } \textit{initially} \quad & -\frac{\partial A_r}{\partial a_r} = J_{rr} - \lambda I_{rr}, \quad -\frac{\partial A_r}{\partial a_s} = J_{rs} - \lambda I_{rs}, \end{aligned}\right\} \quad (55)$$

λ having its value according to (54).

The formulae (54) and (55) give initial values of the "forces" and "influence coefficients". As in § 18, having approximately "liquidated" the residual forces in a first stage, we have interim values of $a_2, \dots, a_n$ which can be used to calculate revised values of λ, of the residual forces, and of the influence coefficients wherewith we can start the second stage. The new value of λ $(= N/D)$ comes from (53); the new values of the residual forces come from

$$\left.\begin{aligned} & -2A_r = D\frac{\partial \lambda}{\partial a_r} = \frac{\partial N}{\partial a_r} - \lambda \frac{\partial D}{\partial a_r}; \\ & \text{and the new values of the influence coefficients come from} \\ & -2\frac{\partial A_r}{\partial a_r} = \frac{\partial^2 N}{\partial a_r^2} - \lambda \frac{\partial^2 D}{\partial a_r^2} + 2\frac{A_r}{D}\frac{\partial D}{\partial a_r}, \\ & -2\frac{\partial A_r}{\partial a_r} = \frac{\partial^2 D}{\partial a_r \partial a_s} - \lambda \frac{\partial^2 D}{\partial a_r \partial a_s} + 2\frac{A_s}{D}\frac{\partial D}{\partial a_r}, \end{aligned}\right\} \quad (29)\,\textit{bis}$$

$$\text{with} \quad \left.\begin{aligned} \frac{1}{2}\frac{\partial N}{\partial a_r} &= J_{1r} + a_2 J_{2r} + \dots + a_n J_{nr}, \\ \frac{1}{2}\frac{\partial D}{\partial a_r} &= I_{1r} + a_2 I_{2r} + \dots + a_n I_{nr}, \end{aligned}\right\} \quad (56)$$

according to (51) and (53). As before the terms involving A_r, A_s in the expressions for the influence coefficients may be suppressed in any one stage, on the ground that A_r, A_s will be negligible at the close of that stage.

28. Relaxation may be continued to any extent which is deemed necessary, and we shall have, finally, a close estimate both to λ_1 and to the associated mode $\bar{Y}_1$. Then, as in §§ 19–21, we can utilize our knowledge of $\bar{Y}_1$ to restrict the mode Y so that it has no component $\bar{Y}_1$, and in this way reduce the finding of λ_2 (or of λ_3 in a symmetrical problem; cf. § 21) to a further *minimal* problem.

The condition corresponding with (41) is

$$\int_0^1 \mu(z)\, Y' \bar{Y}_1'\, dz = -\int_0^1 Y \frac{d}{dz}\{\mu(z)\, \bar{Y}_1'\}\, dz = 0, \tag{57}$$

and we can arrange to satisfy it automatically by the series expression

$$Y = a_2(Y_2 - \alpha_2 Y_1) + a_3(Y_3 - \alpha_3 Y_1) + \ldots + a_n(Y_n - \alpha_n Y_1), \qquad (42)\, bis$$

provided that α_r is determined from

$$\alpha_r \int_0^1 Y_1 \frac{d}{dz}\{\mu(z)\, \bar{Y}_1'\}\, dx = \int_0^1 Y_r \frac{d}{dz}\{\mu(z)\, \bar{Y}_1'\}\, dz, \tag{58}$$

which corresponds with (43). Having determined $\bar{Y}_1$ and $\bar{Y}_2$ we can proceed to determine $\bar{Y}_3$, and so on.

AN EXAMPLE TRACTABLE BY EXACT ANALYSIS

29. It is not easy to devise examples, completely soluble by exact analysis, in which both $\beta(z)$ and $\mu(z)$ are functions of z. The example which follows (suggested by the strut problem of § 8) is artificial in that flexural rigidity increasing from the ends to the centre is associated with a mass which decreases; but it will serve to test the accuracy of our approximate methods.

Let both ends be simply supported, and let

$$\beta(z) = \{1 - 4k^2 z^2\}^2, \quad \mu(z) = \{1 - 4k^2 z^2\}^{-2}, \tag{59}$$

z being measured from the central section. The natural frequencies of vibration are given exactly by

$$\sqrt{\lambda} = 4k^2 \left[1 + \frac{n^2 \pi^2}{\left\{ \log_e \left(\frac{1+k}{1-k} \right) \right\}^2} \right], \tag{60}$$

when λ has the significance given in (50) and n has any integral value. The mode corresponding with each frequency is given by

$$\left. \begin{aligned} Y &\propto (1-4k^2z^2)^{\frac{1}{2}} \cos\left\{\mu \log_e\left(\frac{1+2kz}{1-2kz}\right)\right\} \quad \text{when } n \text{ is odd,} \\ Y &\propto (1-4k^2z^2)^{\frac{1}{2}} \sin\left\{\mu \log_e\left(\frac{1+2kz}{1-2kz}\right)\right\} \quad \text{when } n \text{ is even,} \end{aligned} \right\} \tag{61}$$

μ being a constant defined by

$$4\mu^2 = \frac{\sqrt{\lambda}}{4k^2} - 1. \tag{62}$$

When $k = 0{\cdot}8$ the first four values of λ as calculated from (60) are

$$\lambda_1 = 60{\cdot}74, \quad \lambda_2 = 552{\cdot}0, \quad \lambda_3 = 2466, \quad \lambda_4 = 7447. \tag{63}$$

30. Applied to this example (by K. N. E. B.) a slight modification of the method of "relaxation by inspection" (§§ 6–10) gave an estimate 60·32 for λ_1: the fractional error of this figure is $-0{\cdot}7$ % (i.e. the first natural frequency was underestimated by 0·35 %), and the associated mode was determined with an error everywhere less than 1 %. λ_3 was underestimated by 1·8 % (an error of only 0·9 % in the frequency), but the mode was not so satisfactorily determined, the errors amounting to 8 %. Moreover, the method was not satisfactory as applied to "higher" modes: it seemed as though the rod were trying to assume one of the configurations associated with lower frequencies.

TABLE VIII. MODES AND NATURAL FREQUENCIES OF VIBRATION FOR NON-UNIFORM SHAFT

	First mode		Third mode		Fifth mode	
$z \times 20$	By relaxation method	By exact analysis	By relaxation method	By exact analysis	By relaxation method	By exact analysis
0	1	1	1	1	1	1
1	0·9893	0·9902	0·9346	0·9385	0·8078	0·8393
2	0·9580	0·9613	0·7423	0·7629	0·3235	0·4072
3	0·9059	0·9025	0·4446	0·4780	−0·2568	−0·1711
4	0·8328	0·8429	0·0863	0·1399	−0·7370	−0·6797
5	0·7392	0·7532	−0·2755	−0·2235	−0·9186	−0·9106
6	0·6256	0·6432	−0·5806	−0·5463	−0·6463	−0·7254
7	0·4939	0·5118	−0·7798	−0·7541	−0·0077	−0·1550
8	0·3420	0·3594	−0·7258	−0·7637	0·5545	0·4996
9	0·1755	0·1871	−0·4498	−0·5069	0·5513	0·6807
10	0	0	0	0	0	0
λ	60·96	60·74	2589	2466	19840	17795

31. Consequently the same problem was attacked (by D. G. C.) on the lines which have been indicated in §§ 27–28. It is not necessary to describe the calculations in detail, since they were exactly similar to those of §§ 22–23. On account of symmetry the modes fall into two classes (cf. § 21): attention was confined to "odd" modes (i.e. to λ_1, λ_3, λ_5), and in each instance the series taken as representing the mode was

$$Y = a_1 Y_1 + a_3 Y_3 + a_5 Y_5 + a_7 Y_7 + a_9 Y_9, \tag{64}$$

Y_n having the form $\sin n\pi z$. The device of § 28 was not required.

The integrals of types I_{rs}, J_{rs} are not tabulated, since all (cf. § 27) are special to this particular example. Table VIII summarizes the results, and fig. 4 serves to indicate the approximation to which the modes have been determined.

CONCLUSION

32. In this paper we have chosen for detailed treatment the simplest examples, respectively, of elastic stability and vibrations. But an equation of type (22) can be formulated for any stability problem whether in one or more dimensions (P defining the magnitude of the action making for instability), and equation (47), similarly, has its counterpart in every problem concerning free vibrations. The principles as distinct from the details of our "relaxation" treatment are quite general, and they provide a new method of attack on problems which are at once difficult and important.

The advantages in these problems of "systematic relaxation" as contrasted with "relaxation by inspection" would probably be maintained in relation to configurations of static equilibrium for continuous systems (the problem treated in Part I of this series), although the latter was found to be a satisfactory method of attack.

SUMMARY

Preceding papers in this series have been concerned with unique solutions: this paper deals with problems such that a complete solution would define an infinite number of configurations, all satisfying the governing equations and the terminal conditions, but each associated with some particular value of a critical loading or frequency (the "characteristic number" or *Eigenwert*). The strut and vibrating rod (of non-uniform section) have been chosen for detailed treatment as the simplest examples, respectively, of elastic stability and vibrations; but the principles as distinct from

the details of the relaxation treatment are quite general, and they provide a new method of attack on problems which are at once difficult and important.

In both problems Relaxation Methods appear capable of determining, with more than sufficient accuracy for practical purposes (when dimensions, etc. are not known exactly), not only characteristic numbers but also the mode associated with each. (The modes are not determined by the well-known method of Lord Rayleigh.)

REFERENCES

Atkinson, R. J., Bradfield, K. N. E. and Southwell, R. V. 1937 *Rep. Memor. Aero. Res. Comm., Lond.*, no. 1822.
Black, A. N. and Southwell, R. V. 1938 *Proc. Roy. Soc.* A, **164**, 447-67.
Cowley, W. L. and Levy, H. 1918 *Rep. Memor. Aero. Res. Comm., Lond.*, no. 484.
Southwell, R. V. 1936 "Introduction to the Theory of Elasticity." Oxford Univ. Press.
Temple, G. and Bickley, W. G. 1933 "Rayleigh's Principle." Oxford Univ. Press.

Critical and co-operative phenomena

III. A theory of melting and the structure of liquids

BY J. E. LENNARD-JONES, F.R.S., AND A. F. DEVONSHIRE, PH.D.
The University Chemical Laboratory, Cambridge

(*Received* 29 *August* 1938)

1. INTRODUCTION

There have been many theories of the melting of solids but none has as yet given a satisfactory explanation of the change of phase from solid to liquid in terms of interatomic forces. They have usually been based on simple models and have attempted only to correlate the temperature of melting with other properties of the solid. Of the earlier theories the best known is that of Lindemann (1910), who made the assumption that melting takes place when the amplitude of the vibrations of the atoms in a crystal is a certain fraction of the distance between nearest neighbours. This fraction was determined from the data for one crystal and then applied to others of the same type. This effected a surprisingly good correlation between the melting-points of many solids and their characteristic frequencies.

Raschevsky (1927) and later Herzfeld and Goeppert Mayer (1934) introduced another criterion for melting. The latter showed that if a lattice is expanded uniformly at constant temperature it reaches a state for which the pressure is a minimum with respect to changes of volume. This they associated with melting. Though the theory does not give the heat of fusion, it gives the temperature of melting as a function of pressure. On the other hand, Raschevsky's theory gives the heat of fusion but not the temperature of melting.

An attempt on different lines was made by Braunbek (1926), who considered binary crystals and supposed that the two lattices move rigidly with respect to each other. He assumed their mutual potential energy to be a periodic function of position and showed that there were two possible values for the potential energy for each value of the mean kinetic energy. One of these he attributes to the solid state and the other to the liquid state. By adjusting the amplitude of the potential function he was able to obtain a relation between the heat of fusion and the temperature of melting in fair agreement with experiment.

Frenkel has discussed the theory of fusion in a qualitative way (1937) and has given arguments for supposing that the transition from the solid crystalline state to the liquid one can be carried out in a continuous manner by gradually increasing the volume at constant temperature. He suggests that with increase of volume the regularity will decrease until the assembly becomes amorphous, and discusses the possibility of obtaining an isotherm similar to that of van der Waal's theory of condensing vapours though for quite different reasons and in entirely different temperature regions, the loss of regularity somehow playing an important role. No attempt was made, however, to develop a quantitative theory.

2. ORDER AND DISORDER IN LIQUID AND SOLID

The first difficulty to be resolved in constructing a theory of melting of a normal substance is to find the essential difference between the solid and liquid states. The X-ray examination of liquids has shown that they have regions of order which are almost crystalline, at any rate just above the melting-point, while recent work on the structure of alloys has shown that certain types of disorder can occur in the solid state (Bragg and Williams 1934; Bethe 1935). The difference between a crystalline solid and a liquid can probably be attributed to the disappearance of long-distance order on melting, though some degree of short-distance order may still persist in the liquid and be lost gradually as the temperature rises or the volume increases.

No theory of the gradual decrease of regularity has yet been given, though an interesting suggestion has been made by Eyring and his co-workers (Hirschfelder, Stevenson and Eyring 1937). They suggest that, whereas the atoms in a solid are localized most of the time, those in a liquid can continuously interchange places and can in consequence share the whole of the available volume. This latter property introduces an important change into the statistical treatment, for it changes a factor $(V/N)^N$, which occurs in the partition function of the solid, to $V^N/N!$ and this leads to a new term in the entropy of amount k (Boltzmann's constant), corresponding to the increased disorder of the liquid. There is nothing in the theory, as at present formulated, to indicate what are the conditions for this sudden change from perfect order to complete disorder, nor is there an explanation of the sudden change of volume which is known to occur at the melting-point. A satisfactory theory of melting should lead us to understand more clearly why for a given pressure melting occurs at a definite and precise temperature (and not gradually over a range of temperature) and why it is accompanied by an increase of volume. Eyring's suggestion does, however, focus attention on the essential difference between a solid and a liquid, namely, that a solid at most temperatures can be regarded as *ordered*, while a liquid for the most part is *disordered*.

In this paper we attempt to construct a theory of melting by introducing a theory of disorder. We confine our attention to assemblies of atoms (or molecules), whose fields of force are spherically symmetrical, so that there are no specific points of attachment such as are believed to exist in abnormal liquids like water. Atoms with fields of this type probably form what are called normal liquids, though no strict definition exists as to what a normal liquid is. Usually a liquid is regarded as normal when it obeys certain empirical rules such as Trouton's rule. This latter rule was derived in terms of interatomic forces of the spherical type in a recent paper (Lennard-Jones and Devonshire, Paper II, 1937) and probably implies non-directed fields of force.

We take over the theory and methods of the two preceding papers (Lennard-Jones and Devonshire 1937, under the title of "Critical Phenomena in Gases I and II"), according to which each atom of a dense assembly, whether solid, liquid or gas, is regarded as moving in the field of force of its neighbours and confined by them to a small region of space equal to the average available volume per atom. This method proved surprisingly successful in providing direct calculations of critical temperatures and also of boiling-points in terms of interatomic forces. But the model as thus described does not differentiate a solid from a liquid. In the solid the cell in which an atom moves is localized whereas in a liquid it is not. There is a migration of the cell from one position to another. In a liquid therefore we must represent the motion of an atom as a vibration about a point which itself suffers displacement. This feature we introduce into the theory in this paper by considering the order and disorder of the sites (or centres of vibration) and discuss the relation between solid and liquid in terms of the disorder present.

We picture a solid of one kind of atom (such as that of an inert gas) as an alloy of atoms and "holes", the atoms occupying at low temperatures all the available sites on a perfect lattice and the holes all the available sites on a similar lattice interpenetrating the first. A close analogy in two dimensions would be a draught or chess board, of which the white squares constitute the normal sites for the atoms and the black squares abnormal sites. We call the former α-sites and the latter β-sites. Each site is surrounded by z sites of the other type. When an atom is situated at an α-site and another at a contiguous β-site, we shall call them *close* neighbours, the term *immediate* neighbours being reserved, if need be, for the nearest sites of the same type.

A measure of the disorder in such an assembly will be the number of atoms in abnormal sites. Fig. 1*a* shows a two-dimensional arrangement of α- and β-sites. Most of the atoms are ordered, but one or two are in β-sites. This arrangement corresponds to a state of long-distance order but local regions of disorder. A state of complete disorder is shown in fig. 1*b*, where the atoms have been placed at random on α- and β-sites in such a way that there are equal numbers on each. In such an arrangement all evidence of a regular lattice structure has disappeared. We do not suggest that the actual configuration, shown in fig. 1*b*, is a likely one. The clusters will not remain congested while surrounded by empty sites, but will expand until they occupy as much as possible of the available volume. This aspect of the problem we shall consider later. What we want to do first is to get a mathematical theory of disorder to give the *relative* positions of the atoms surrounding any one, and then we shall be able to deal later with the problem as to what will be their actual positions.

The theory of disorder in alloys consisting of two kinds of atoms has been considered by several authors and has been successful in accounting for anomalous properties of these alloys at a certain temperature, called the critical temperature (Bragg and Williams 1934–5; Bethe 1935; Williams 1935). This temperature is now interpreted as the one at which long-distance order disappears. In this paper we shall consider a solid of one kind of atom as an alloy of atoms and holes, and the migration of an atom to an abnormal site as the interchange of an atom and a hole. We adopt with suitable modifications the method used by Bethe for alloys, but we take the energy necessary for an interchange of atom and hole to be a function of the volume of the whole assembly. This makes the phenomenon a co-operative one in a twofold sense and, as we shall see, the combination of these two concepts leads to a theory of melting.

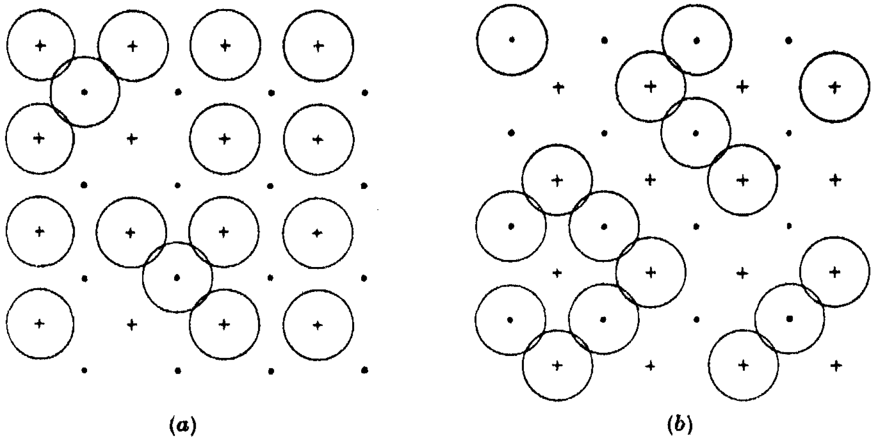


FIG. 1*a*. Local displacements but long-distance order. (α-sites are represented by crosses, β-sites by dots and atoms by circles.)

FIG. 1*b*. Break up of long-distance order, equal numbers on α- and β-sites.

3. THE FREE ENERGY AND PRESSURE DUE TO DISORDER

We suppose that the crystal is in equilibrium with its vapour, the pressure of which is supposed known, and we use the partition function of an atom in the vapour phase as a standard of reference. We denote by f the ratio in which the partition function of the whole system (crystal and vapour) is altered when an atom is transferred from the vapour to an α-site with no close neighbours. The function f is to be determined by the condition that half of the total number of α- and β-sites is occupied. A somewhat similar procedure is used in the theory of adsorption on a surface, when there is co-operation among the adsorbed atoms, and the vapour pressure is known but the number of occupied sites on the surface is unknown (Peierls 1936); in this case the function f is determined so as to give a known vapour pressure.

When an atom is transferred from the vapour to an α-site and has m close neighbours, we shall suppose the partition function changed by a factor $f\exp(-mW/kT)$, where W is a measure of the change of energy required to transfer an atom from a position as an immediate neighbour on the same lattice to one in which it is a close neighbour. The effect of disorder in the crystal on the partition function is introduced, as in Bethe's method, by supposing that the partition function of an atom in one of the β-sites, which is nearest to a central α-site, is altered by a factor ξ owing to the distribution in the rest of the crystal. Similarly, the partition function of an atom in one of the α-sites, which immediately surround a central β-site, is altered by some other unknown factor ξ'. There are then four unknowns, viz. f, ξ, ξ' and the probability (Q) of occupation of an α-site. We now proceed to determine each of these quantities, using the fact that the probability of occupation of each α-site must be the same (Q) and the probability of occupation of each β-site must be $(1-Q)$.

The partition function for an atom on an α-site and m close neighbours is then

$$R_m = \frac{z!}{m!\,(z-m)!}(\eta\xi)^m(\lambda f)^{m+1}, \tag{3·01}$$

where we have written

$$\eta = e^{-W/kT}, \tag{3·02}$$

and λ is the partition function in the vapour phase. The probability of this configuration relative to the configuration in which the α-site and the surrounding sites are all vacant and the $m+1$ atoms are in the gas phase is (3·01) divided by λ^{m+1}, which we may denote by r_m. Summing for all possible m, we then get

$$r_m = R_m/\lambda^{m+1}, \tag{3·03}$$

$$r = \sum_{m=0}^{m=z} r_m = f(1+\eta f\xi)^z. \tag{3·04}$$

Similarly, the partition function for the group when the central α-site is not occupied is

$$r_0 = (1+f\xi)^z. \tag{3·05}$$

The probability that an α-site is occupied then is

$$\begin{aligned} Q &= \frac{f(1+\eta f\xi)^z}{(1+f\xi)^z+f(1+\eta f\xi)^z} \\ &= \frac{f(1+\eta\epsilon)^z}{(1+\epsilon)^z+f(1+\eta\epsilon)^z}, \end{aligned} \tag{3·06}$$

on writing
$$\epsilon = f\xi. \tag{3·07}$$

Now the probability that any one of the β-sites is occupied when the α-site is occupied is equal to

$$\frac{\sum_{m=0}^{z}\frac{m}{z}r_m}{\Sigma r_m} = \frac{\epsilon\frac{\partial}{\partial\epsilon}(1+\eta\epsilon)^z f}{z(1+\eta\epsilon)^z f} = \frac{\eta\epsilon}{1+\eta\epsilon}, \tag{3·08}$$

and similarly the probability that it is occupied when the centre is unoccupied is $\epsilon/(1+\epsilon)$. But the average probability that a β-site is occupied is $1-Q$, and so we have

$$(1-Q) = \left(\frac{\eta\epsilon}{1+\eta\epsilon}\right)Q + \left(\frac{\epsilon}{1+\epsilon}\right)(1-Q), \tag{3·09}$$

giving the relation
$$\frac{Q}{1-Q} = \frac{(1+\eta\epsilon)}{\eta\epsilon(1+\epsilon)}, \tag{3·10}$$

and
$$= \frac{N_\alpha}{N_\beta}, \tag{3·11}$$

where N_α and N_β are the numbers on α- and β-sites respectively. Similarly, by considering a β-site as central we get another set of equations, which may be written

$$1-Q = \frac{f(1+\eta\epsilon')^z}{(1+\epsilon')^z + f(1+\eta\epsilon')^z}, \tag{3·12}$$

and
$$Q(1+\eta\epsilon') = (1-Q)\,\eta\epsilon'(1+\epsilon'). \tag{3·13}$$

Equations (3·06), (3·10), (3·12) and (3·13) give four relations between the four unknowns ϵ, ϵ', Q and f, which may be rewritten in the form

$$\frac{Q}{1-Q} = \frac{f(1+\eta\epsilon)^z}{(1+\epsilon)^z} = \frac{(1+\epsilon')^z}{f(1+\eta\epsilon')^z}$$
$$= \frac{(1+\eta\epsilon)}{\eta\epsilon(1+\epsilon)} = \frac{\eta\epsilon'(1+\epsilon')}{1+\eta\epsilon'}. \tag{3·14}$$

These equations determine Q, f, ξ and ξ' in terms of η, that is in terms of W/kT.

Eliminating Q and f we obtain

$$(1+\eta\epsilon)^{z-1}(1+\epsilon)^{-(z-1)}\epsilon = (1+\eta\epsilon')^{z-1}(1+\epsilon')^{-(z-1)}\epsilon', \tag{3·15}$$

and
$$(1+\eta\epsilon)(1+\eta\epsilon') = \eta\epsilon\eta\epsilon'(1+\epsilon)(1+\epsilon'), \tag{3·16}$$

that is
$$(1-\eta\epsilon\epsilon')\{1+\eta(\epsilon+\epsilon'+\epsilon\epsilon')\} = 0, \tag{3·17}$$

so that
$$\eta\epsilon\epsilon' = 1, \tag{3·18}$$

since η, ϵ and ϵ' are always positive. Substituting for ϵ' in equation (3·15), we find, on writing

$$x = \eta^{\frac{1}{2}}\epsilon, \tag{3·19}$$

an equation to determine x and so ϵ, viz.

$$(1+\eta^{\frac{1}{2}}x)\,x^{1/(z-1)} = \eta^{\frac{1}{2}}+x. \tag{3·20}$$

This equation always has a root given by $x=1$, and if the left-hand side and the right-hand side be plotted separately as functions of x the two curves intersect in two other points, provided that

$$z\eta^{\frac{1}{2}} < (z-2). \tag{3·21}$$

The root $x=1$ corresponds to $Q=\frac{1}{2}$ and no long-distance order. As for the other roots, we note that if x_1 is a root of equation (3·20), then $1/x_1$ is also a root, and further we have from equation (3·14) for a root

$$\frac{Q}{1-Q} = \left(\frac{1}{x_1}\right)^{z/(z-1)}, \tag{3·22}$$

so that the two roots, which are not equal to unity, are equivalent to one another, the value of Q corresponding to one root being equal to the value of $(1-Q)$ for the other. Both correspond to a state of long-distance order and it may be shown that where they exist they imply a minimum value of the free energy, that is a state of stability. Since they are equivalent, it will be sufficient to take the lowest root only, which corresponds to a value of Q between $\frac{1}{2}$ and 1, or a preponderance of atoms on α-sites.

The relation

$$\exp(-W/2kT_c) = \eta^{\frac{1}{2}} = (z-2)/z, \tag{3·23}$$

defines a critical temperature T_c at which order disappears. In this paper we shall consider W to be a function of volume, as it certainly must be, and so, if the crystal be heated at constant volume, the temperature at which order disappears is a function of the volume. This step is new and the crucial one, as we shall see, in the theory of melting.

The additional energy of the system due to disorder is given by*

$$U'' = \frac{NzWQ\eta\epsilon}{(1+\eta\epsilon)} = \frac{NzWQ\eta^{\frac{1}{2}}x}{(1+\eta^{\frac{1}{2}}x)}, \tag{3·24}$$

---

* We denote by U'' the additional internal energy of the assembly as a result of disorder, reserving U' to denote the internal energy of the assembly when ordered. Similarly we denote the additional entropy, free energy and pressure by S'', A'' and p''.

since each atom in an α-site has on the average $z\eta\epsilon/(1+\eta\epsilon)$ neighbours on close β-sites, and there are NQ such atoms. Substituting from (3·14), we have

$$U'' = \frac{NzW\eta^{\frac{1}{2}}x}{(1+2\eta^{\frac{1}{2}}x+x^2)}, \tag{3·25}$$

as long as there is any long-distance order in the crystal. For $z\eta^{\frac{1}{2}} \geqslant (z-2)$ we have long distance disorder and $x=1$, so that

$$U'' = NzW\eta^{\frac{1}{2}}/2(1+\eta^{\frac{1}{2}}). \tag{3·26}$$

For complete disorder we have $\eta = 1$ and $U'' = \frac{1}{4}NzW$.

The corresponding contribution to the Helmholtz free energy can be obtained from the relation

$$A'' = -T\int \frac{U''}{T^2}dT = -kT\int \frac{U''}{\eta W}d\eta$$

$$= -NkTz\int_0^\eta \frac{\eta^{-\frac{1}{2}}x\,d\eta}{1+2\eta^{\frac{1}{2}}x+x^2}, \tag{3·27}$$

as long as $z\eta^{\frac{1}{2}} < (z-2)$, or

$$T < (W/2k)\,[\log z/(z-2)]^{-1}. \tag{3·28}$$

For temperatures above the critical temperature, we have

$$A'' = -T\int_0^{T_c} \frac{U''}{T^2}dT - T\int_{T_c}^{T} \frac{U''}{T^2}dT$$

$$= -NkTz\int_0^{\eta_c} \frac{\eta^{-\frac{1}{2}}x\,d\eta}{1+2\eta^{\frac{1}{2}}x+x^2} - NkTz\int_{\eta_c}^{\eta} \frac{\eta^{-\frac{1}{2}}d\eta}{2(1+\eta^{\frac{1}{2}})}, \tag{3·29}$$

where

$$\eta_c = (z-2)^2 z^{-2}. \tag{3·30}$$

Now, if we substitute for $\eta^{\frac{1}{2}}$ in the first integral from equation (3·20), we have

$$\eta^{\frac{1}{2}} = \frac{x^{1/(z-1)} - x}{1 - x^{z/(z-1)}} \tag{3·31}$$

and

$$z\int_0^\eta \frac{\eta^{-\frac{1}{2}}x\,d\eta}{1+2\eta^{\frac{1}{2}}x+x^2} = 2z\int_0^x \frac{x(-1+x^{2/(z-1)}) + x^{1/(z-1)}(1-x^2)/(z-1)}{(1-x^2)(1-x^{2z/(z-1)})}\,dx$$

$$= 2z\int_0^x \left\{-\frac{x}{1-x^2} + \frac{x^{(z+1)/(z-1)}}{1-x^{2z/(z-1)}}\right\}dx + \frac{z}{(z-1)}\int_0^x \left\{\frac{x^{1/(z-1)}}{1+x^{z/(z-1)}} + \frac{x^{1/(z-1)}}{1-x^{z/(z-1)}}\right\}dx.$$

Hence for temperatures below the critical temperature

$$-A''/NkT = z\log(1-x^2)-(z-1)\log(1-x^{2z/(z-1)}) \\ +\log(1+x^{z/(z-1)})-\log(1-x^{z/(z-1)}). \quad (3{\cdot}32)$$

When the upper limit is η_c or $x=1$, this becomes

$$z\log\{(z-1)/z\}+2\log 2. \quad (3{\cdot}33)$$

The second integral in (3·29) gives

$$z\log\{(1+\eta^{\frac{1}{2}})/2\}-z\log\{(z-1)/z\}, \quad (3{\cdot}34)$$

and so for temperatures above T_c, we get

$$-A''/NkT = 2\log 2+z\log\{(1+\eta^{\frac{1}{2}})/2\}. \quad (3{\cdot}35)$$

The limiting value of A'' as W tends to zero (infinite specific volume) or the temperature tends to infinity is

$$-A''/NkT = 2\log_e 2, \quad (3{\cdot}36)$$

as it should be, for then it is due to the extra entropy corresponding to the shuffling of N atoms among $2N$-sites.

The variation of U'', A'' and the entropy S'' with volume for a constant temperature is shown for a typical case in Fig. 3.

The additional pressure due to the change of order is given by

$$p'' = -\left(\frac{\partial A''}{\partial v}\right)_T,$$

and, as (A''/NkT) is a function of η only, we have

$$p'' = -NkT\left[\frac{\partial}{\partial v}\left(\frac{A''}{NkT}\right)\right]_T = -NkT\left(\frac{\partial\eta}{\partial v}\right)_T\frac{d}{d\eta}\left(\frac{A''}{NkT}\right) \\ = -\frac{1}{W}\left(\frac{dW}{dv}\right)U'', \quad (3{\cdot}37)$$

on using (3·02) and (3·27). We thus have the following explicit formulae for p'', using (3·24) and (3·25), viz.

$$p'' = -Nz\frac{dW}{dv}\frac{\eta^{\frac{1}{2}}x}{1+2\eta^{\frac{1}{2}}x+x^2} \quad (3{\cdot}38)$$

for $\eta^{\frac{1}{2}} < (z-2)/z$, and $p'' = -Nz\dfrac{dW}{dv}\dfrac{\eta^{\frac{1}{2}}}{2(1+\eta^{\frac{1}{2}})}$ (3·39)

for $\eta^{\frac{1}{2}} \geqslant (z-2)/z$. We note that these formulae can also be written in the form

$$p'' = -n''\frac{dW}{dv}, \quad (3{\cdot}40)$$

where n'' is the average number of pairs of close neighbours at any given temperature and volume.

Further, if W depends on volume according to an inverse power law, so that it can be written in the form

$$W = W_0(v_0/v)^l, \tag{3·41}$$

we have the simple formula for p'',

$$p'' = lU''/v. \tag{3·42}$$

This shows that p'' has a maximum. For at a given temperature U'' vanishes at small volumes (and more quickly than v) and tends to $\frac{1}{4}NzW$ at large volumes, so that it again vanishes in virtue of the diminishing W. A typical curve for p'' is shown in fig. 2.

The formula (3·40) shows that for a given temperature p'' will reach a sharp maximum at a volume approximately equal to that at which n'' reaches its maximum constant value, that is at the volume for which long-range order disappears. This is determined by

$$W_c = kT\log\{z/(z-2)\}, \tag{3·43}$$

which gives a critical energy W_c for the given temperature and hence a critical volume. Moreover, for this volume and temperature, we find from (3·26) that

$$U''_c = \tfrac{1}{4}NzW(z-2)/(z-1), \tag{3·44}$$

so that, using (3·42),

$$\begin{aligned} p''_c &= l(NzW_c/4)\{(z-2)/(z-1)\}/v_c \\ &= \frac{lz(z-2)\log\{z/(z-2)\}}{4(z-1)}\,\frac{NkT}{v_c}; \end{aligned} \tag{3·45}$$

p''_c is, in fact, nearly equal to the maximum value of p'', as is evident in fig. 2.

From (3·43) we have

$$v_c = v_0\left[\frac{W_0}{kT\log\{z/(z-2)\}}\right]^{1/l}, \tag{3·46}$$

and hence we see that p'' at this critical volume is proportional to $T^{1+(1/l)}$; in fact, we have

$$p''_c = lA(z)\frac{NkT}{v_0}\left(\frac{kT}{W_0}\right)^{1/l}, \tag{3·47}$$

where $A(z)$ is a function of z only and the interatomic forces enter through W_0, v_0 and l.

4. THE EQUATION OF STATE OF SOLID AND LIQUID

Having found the pressure and free energy due to disorder, it is now necessary to consider the corresponding values of these quantities due to ordinary heat motion in an ordered assembly. Now this is precisely what we have calculated in earlier papers (Lennard-Jones 1937; Lennard-Jones and Devonshire, Paper I, 1937). We there supposed each atom to move in a cell, determined by its neighbours, and calculated the potential energy for all positions within the cell, supposing all the other atoms located at the centres of their cells. This field was then averaged so as to give it spherical symmetry and we defined a free or available volume per atom as

$$v_f^* = \int \exp\{-\Delta\psi(r_1 v)/kT\}\, 4\pi r^2\, dr, \tag{4·01}$$

an integral over the cell, $\Delta\psi$ being the potential energy difference between any point of the cell and that at the centre due to the atoms in the immediate neighbourhood. In constructing the partition function of the whole assembly, we supposed that as a first approximation the atoms might be regarded as moving independently of each other, as in Einstein's theory of the specific heat of solids. In that case the partition function could be written

$$F(v, T) = [f(v, T)]^N, \tag{4·02}$$

where $$f(v, T) = (2\pi mkT/h^2)^{\frac{3}{2}}\, v_f^* \exp(-\Phi_0(v)/NkT), \tag{4·02 A}$$

and Φ_0 is the mutual potential energy of the assembly when each is at the centre of its cell (Lennard-Jones 1937, equation (16)). This partition function is just of the type we require, for it is valid for a set of atoms which are localized about the points of a regular lattice, the α-sites, in fact, of the preceding work. It is not to be multiplied by any factor such as $N^N/N!$, as was done in Paper II in the theory of a liquid. The factor by which the partition function is to be multiplied to take account of disorder is implicit in the discussion of the preceding section.

The free energy appropriate to (4·02) is then given by

$$\begin{aligned} A' &= -NkT \log f(v, T) \\ &= -NkT \log\left(\frac{2\pi mkT}{h^2}\right)^{\frac{3}{2}} - NkT \log v_f^* + \Phi_0(v), \end{aligned} \tag{4·03}$$

and the pressure by $$p' = \frac{NkT}{v_f^*}\left(\frac{\partial v_f^*}{\partial v}\right)_T - \frac{\partial \Phi_0}{\partial v}. \tag{4·04}$$

Actual calculations were made of the available volume v_f as a function of the volume of the assembly and of the temperature, assuming the potential of the interatomic field of two atoms to be expressed as a sum of two inverse power laws. This type of field can be written as

$$\phi = -|\phi_0|\left\{\frac{1}{n}\left(\frac{r_0}{r}\right)^n - \frac{1}{m}\left(\frac{r_0}{r}\right)^m\right\}\Big/\left(\frac{1}{n}-\frac{1}{m}\right), \tag{4·05}$$

and then $|\phi_0|$ is the energy required to separate two atoms from their position of minimum energy to infinity, r_0 is their distance apart in the minimum configuration, and n, m are suitable indices of the repulsive and attractive fields. Values of n and m were chosen which have been found successful in the theory of the virial coefficient (Lennard-Jones 1937), viz. $n = 12$, $m = 6$.

The atoms were assumed to be arranged on a face centred lattice, the average distance between nearest neighbours being a, and the specific volume then being $v^* = a^3/\sqrt{2}$. Inserting the above values of n and m and putting $v_0^* = r_0^3/\sqrt{2}$, the potential energy function (4·05) can be rewritten as

$$\phi = |\phi_0|\{(v_0^*/v^*)^4 - 2(v_0^*/v^*)^2\}. \tag{4·06}$$

In this notation the expression for p' given in Paper I, equation (50) becomes

$$p' = \frac{kT}{v^*}\left[1 - \frac{c|\phi_0|}{kT}\left\{(2.4)\left(\frac{v_0^*}{v^*}\right)^2 - 2\left(\frac{v_0^*}{v^*}\right)^4\right\} - \frac{4c|\phi_0|}{kT}\left\{\left(\frac{v_0^*}{v^*}\right)^2\frac{g_m}{g} - \left(\frac{v_0^*}{v^*}\right)^4\frac{g_l}{g}\right\}\right], \tag{4·07}$$

where $c = 12$ and is the number of nearest neighbours. The g's are functions of v and T given explicitly in Paper I, equations (49), (51) and (52), and the symbol A, there used, is equal to $c|\phi_0|$. In terms of the g's the free energy A' is given by equation (4·03) above, using the relation

$$v_f^*/v^* = 2\pi(\sqrt{2})g. \tag{4·08}$$

Now the previous treatment of disorder implicitly associated a state of equivalence between α- and β-sites except that the simultaneous occupation of contiguous sites meant a state of higher energy. The question now arises as to whether the total pressure is given by

$$p = p' + p'', \tag{4·09}$$

for this would imply that each atom has the same available volume v_f^* whether it is situated on an α- or a β-site irrespective of its neighbours. At first sight this would not appear to be likely. An examination of fig. 1*b* is, however, instructive in this connexion. Whenever the disorder leads to a close clustering of groups, it is to be observed that each is surrounded by a series of unoccupied sites. Each cluster would then in actual fact expand under the influence of the atomic forces until the energy is as small as possible, that is until the cluster touches other surrounding clusters. This implies local expansions of the lattice to which the atoms are referred. Each atom will then occupy on the average the same available volume as before, if the total volume of the assembly is kept constant. We consider therefore that it will be sufficiently accurate for the model to take the free energy of the assembly to be given by

$$A = A' + A'', \tag{4·10}$$

and the pressure by (4·09).

The object of a theory of melting must be to calculate the temperature of melting, the change of volume at melting and the latent heat directly from the interatomic forces. To complete the present theory it is necessary therefore to try to find W, the extra energy of a pair of atoms on neighbouring α- and β-sites, in terms of interatomic forces. We have made an estimate of this quantity in the following way. According to our model, perfect order corresponds to a perfect face centred lattice, while complete disorder corresponds to an equal probability of finding an atom on any of the sites of simple cubic lattice. (Actually the face centred lattice is the cubic lattice of lowest potential energy for a field of force of the type given above in (4·05) (Lennard-Jones and Ingham 1925).) We may describe disorder, according to this theory, as a *tendency* for the atoms to set in clusters with a symmetry of the simple cubic lattice. Now, as explained above, the clusters will also tend to expand until each atom has the same atomic volume as before. There may be many plausible ways of expressing the net effect of both these tendencies. One is to suppose tentatively that each cluster will tend to set according to a simple cubic lattice with *every* site occupied (that is, with a probability unity instead of having a probability of only half), but with a spacing such that the potential energy of each cluster is a minimum for this configuration. The latter condition will imply local expansion.

The potential energy of the central atom will then be

$$\psi_s = |\phi_0| \Sigma\left\{\left(\frac{r_0}{d}\right)^{12} - 2\left(\frac{r_0}{d}\right)^{6}\right\} + \Delta\phi, \tag{4·11}$$

for the field already discussed above, d being the distance between atoms. The summation is over the atoms of the cluster (say s in number) and $\Delta\phi$ represents the effect of all the atoms outside it. We could calculate $\Delta\phi$ by replacing all the atoms outside the cluster by a continuum, but it will be sufficiently accurate to regard these atoms also as part of the same lattice. We then get

$$\psi_s = |\phi_0| \left\{ A_{12}\left(\frac{r_0}{d_0}\right)^{12} - 2A_6\left(\frac{r_0}{d_0}\right)^{6} \right\}, \tag{4·12}$$

where d_0 is the spacing of the lattice, while A_6 and A_{12} are summations over an infinite simple cubic lattice and have been given by Lennard-Jones and Ingham (1925). The atomic volume is equal to d_0^3, where d_0 is the value which makes ψ_s a minimum. This is given by

$$d_0 = r_0(A_{12}/A_6)^{\frac{1}{6}} \tag{4·13}$$

and

$$(\psi_s)_{\min} = -|\phi_0|(A_6^2/A_{12}). \tag{4·14}$$

The potential energy of one atom of a face centred cubic due to all the rest is similarly

$$\Psi_s = |\phi_0| \left\{ C_{12}\left(\frac{r_0}{s_0}\right)^{12} - 2C_6\left(\frac{r_0}{s_0}\right)^{6} \right\}. \tag{4·15}$$

For the minimum configuration we have

$$s_0 = r_0(C_{12}/C_6)^{\frac{1}{6}} \tag{4·16}$$

and

$$(\Psi_s)_{\min} = -|\phi_0|(C_6^2/C_{12}). \tag{4·17}$$

The mutual energy of N atoms in the two configurations is $\frac{1}{2}N\Psi_s$ and $\frac{1}{2}N\psi_s$, and as the total energy increment in the disordered state was given by $U'' = \frac{1}{2}NzW$ (cf. eqn. (3·26)), we have, putting $z = 6$,

$$\begin{aligned} W &= \tfrac{1}{3}[(\psi_s)_{\min} - (\Psi_s)_{\min}] \\ &= \tfrac{1}{3}|\phi_0|[(C_6^2/C_{12}) - (A_6^2/A_{12})] \\ &= 2|\phi_0|, \end{aligned} \tag{4·18}$$

very nearly. The percentage change of volume, deduced from (4·16) and (4·13), is 33 %.

This treatment is, of course, only approximate, for it presupposes a complete change of crystal structure and this at the absolute zero of temperature. It is intended only to give an order of magnitude of W, or rather of the value of W when complete disorder has set in with its consequent change of volume. It is equivalent to replacing the change from solid to liquid by a change of crystal structure, and is a cruder picture than the mathematical theory of disorder represents. The actual value of W will be a function of volume and will presumably be less than that given above as each cluster will accommodate itself to its environment in the best possible way. For example, a change to a body centred cubic would require less energy than that given above, being apparently only (0·27) $|\phi_0|$. We may thus, as a result of this discussion, anticipate that W will be of the same order as $|\phi_0|$.

In order to make numerical calculations we have supposed that W will be determined mainly by the repulsive part of the atomic forces and we have accordingly put

$$W = W_0\left(\frac{v_0}{v}\right)^l, \tag{4·19}$$

and have regarded W_0 as an adjustable parameter, to be determined by one observable, which may be either the temperature of melting, or the change of volume on melting or the latent heat or entropy of fusion. The theory should then give the others. We have actually chosen W_0 to give the observed melting-point. This has to be done by a process of trial and error. In the case of argon, for which the temperature of melting at atmospheric pressure is 83·8° abs., we have $|\phi_0|/kT_m = 1·44$ ($|\phi_0|$ being known for argon from interatomic forces (Lennard-Jones 1937, Table III)). The value of W_0 chosen was such that $W_0/kT_m = 1·505$. Actually we find that a pressure of one atmosphere makes so little difference to the temperature of melting that we have taken the pressure to be zero. Thus we find $W_0 = 1·05\phi_0$ which is of the same order as that found above.

A typical set of results is given in Table I. The terms arising from the partition function without disorder are indicated by a single dash, those from disorder by a double dash, and the resultant terms without a dash. The volume at which the long-range order vanishes is indicated by the figures in italics, cf. eqn. (3·46).

TABLE I. THE PRESSURE AND FREE ENERGY OF SOLID-LIQUID ARGON AS A FUNCTION OF VOLUME ($T = 83·8$ ABS.)

$\frac{v}{v_0}$	$\frac{v_0^2}{v^2}$	$\frac{p'v_0}{kT}$	$\frac{p''v_0}{kT}$	$\frac{pv_0}{kT}$	$\frac{-A'}{NkT}$	$\frac{-A''}{NkT}$	$\frac{-A}{NkT}$	$\frac{S'}{Nk}$	$\frac{S''}{Nk}$	$\frac{S}{Nk}$
1·0259	0·95	0·18	0·71	0·89	12·578	0·039	12·617	3·513	0·221	3·734
1·0540	0·90	−0·89	1·03	0·14	12·567	0·065	12·632	3·712	0·335	4·047
1·0691	0·875	−1·31	1·22	−0·09	12·550	0·082	12·632	3·818	0·409	4·227
1·0846	0·85	−1·67	1·47	−0·20	12·526	0·102	12·628	3·922	0·501	4·423
1·1180	0·80	−2·29	2·08	−0·21	12·460	0·161	12·621	4·138	0·741	4·879
1·1548	0·75	−2·73	2·98	0·25	12·367	0·255	12·622	4·373	1·115	5·488
1·1676	*0·7335*	*−2·82*	*3·33*	*0·49*	*12·331*	*0·294*	*12·625*	*4·451*	*1·266*	5·717
1·1951	0·70	−2·98	3·03	0·05	12·251	0·381	12·632	4·607	1·286	5·893

These results are also shown figs. 2, 3 and 4, the volume at which long distance order disappears being indicated by the dotted line.

The relation between the total pressure and volume for a fixed temperature has a shape similar to that shown in fig. 4, but the minimum and maximum points on this isotherm do not correspond to states of minimum free energy. The same arguments, which are used for the liquid-vapour change of phase, can be applied here with the result that part of the curve must be replaced by a straight line. (In fig. 4 this part coincides with the axis of zero pressure.) At the left-hand point on this straight portion the value of p'' is small, and though the order is not perfect there is some long-range order, but as the volume changes along this straight portion the order-disorder phenomenon sets in and by the time the pure liquid phase is reached all long-distance order has disappeared. There is, however, some degree of short-distance order left and this only disappears at higher volumes (or at higher temperatures than the melting one). This result is in accord with prevalent ideas as to the characteristic structures of solid and liquid.

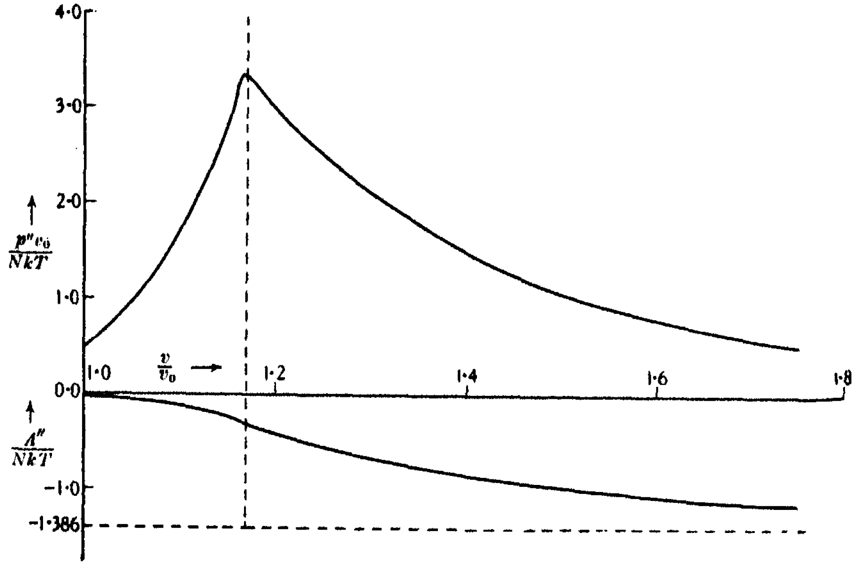


FIG. 2. The pressure and free energy of argon due to disorder in the crystal-liquid change of phase. (The dotted vertical line gives the volume at which long-distance order disappears.)

For the isotherm given in Table I the volume of the solid just before melting is the smallest volume at which the pressure vanishes, namely $1{\cdot}0635v_0$, and the volume of the liquid the largest volume at which the pressure vanishes, namely $1{\cdot}199v_0$. Since the free energies at these two volumes are both equal (A/NkT being equal to $-12{\cdot}632$), the two phases can co-exist and therefore the temperature is the melting temperature at zero pressure; the melting temperature at atmospheric pressure will not differ appreciably from this.

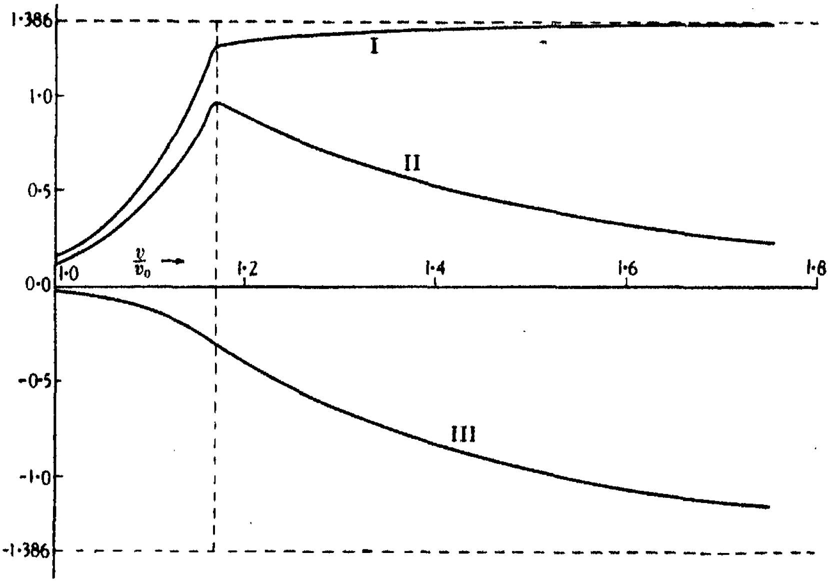


Fig. 3. The entropy S''/Nk (curve I), the energy U''/NkT (curve II) and the free energy A''/NkT (curve III) of argon as a function of volume. (The volume at which long-distance order disappears is shown.)

Table II. The change of volume and entropy of argon and nitrogen at melting

			Change in volume		Entropy of fusion (cal./deg.)		α (thermal expansion of liquid)	
	T_m	W_0/ϕ_0	Calc. %	Obs. %	Calc.	Obs.	Calc.	Obs.
Argon	83·8	1·048	13	12	3·45	3·35	0·0049	0·0045
Nitrogen	63·2	0·875	8	9	2·58	2·73	0·0048	0·0059*

* For the range 68–89°.

From the table we find the entropies corresponding to the volumes of the solid and liquid, and their difference gives the entropy of fusion. The results are given in Table II for argon, together with the ratio of W_0 to ϕ_0, the mutual potential energy of two molecules at their equilibrium distance.

The results of similar calculations for nitrogen are also included. We note that here also W_0 is equal approximately to ϕ_0, which is in accord with the preceding discussion.

The sharp rise in the pressure due to disorder is shown in fig. 2, where the free energy is also given as a function of volume. These curves refer to argon for the temperature at which it melts (at zero pressure). The corresponding values of the entropy (S''/Nk) and energy (U''/NkT) given in fig. 3. In fig. 4 the total pressure is plotted as a function of volume for the melting temperature of argon. The zero axis of pressure replaces the dip and rise in the isotherm. The isotherm without disorder is shown by the dotted curve.

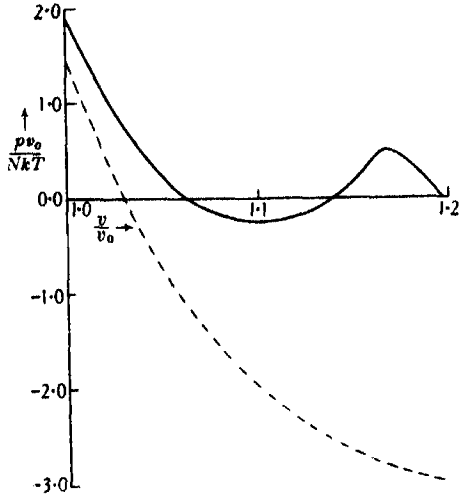


FIG. 4. The total pressure (including that due to disorder) as a function of volume.

Having found W_0 we can now construct an isotherm for higher temperatures, and determine the corresponding pressure of melting. In Table III we give an isotherm for argon corresponding to $|\phi_0|/kT = 4/3$ (90·3° K). It is found by a trial and error process that the pressure of melting then is given by $pv_0/NkT = 0{\cdot}96$. For at this pressure the volume of the solid is given by $v_s = 1{\cdot}0582v_0$ and of the liquid by $v_l = 1{\cdot}1665v_0$, and the corresponding free energies by $A_s = -11{\cdot}989NkT$ and $A_l = -12{\cdot}094NkT$. These quantities satisfy the relation

$$A_s - A_l = p(v_l - v_s),$$

which is the condition for the equilibrium of the two phases. In c.g.s. units the values found for the pressures of melting are for argon 291×10^6

dynes/cm.^{2} at 90·3°, and for nitrogen 171×10^6 at 67·4°, as compared with experimental values of 290×10^6 and 194×10^6 respectively (Simon, Ruhemann and Edwards 1930).

TABLE III. AN ISOTHERM FOR ARGON (CALCULATED) FOR A TEMPERATURE OF 90·3° ABS.

$\dfrac{v}{v_0}$	$\dfrac{v_0^2}{v^2}$	$\dfrac{-A}{NkT}$	$\dfrac{pv_0}{NkT}$	$\dfrac{S}{k}$
1·0259	0·95	11·951	1·62	4·007
1·0540	0·90	11·985	1·02	4·339
1·0846	0·85	12·012	0·79	4·754
1·1180	0·80	12·039	0·94	5·519
1·1463	*0·7614*	*12·072*	*1·38*	*5·767*
1·1548	0·75	12·083	1·20	5·825
1·1951	0·70	12·117	0·49	6·087
1·2404	0·65	12·127	−0·03	6·367

From the table we can also deduce the volume of the liquid at zero pressure, which is $1{\cdot}237v_0$ and hence its average coefficient of thermal expansion (α) over the range 83·8–90·3°. This is given in Table II.

It will thus be seen that when the one empirical constant W_0 has been adjusted to give the right temperature of melting, the other properties associated with the change are in good agreement with experiment. The value of W_0 thus found is less than that given by equation (4·18), but it is of the same order of magnitude and suggests that the treatment of the theory of melting given in this paper is at any rate on the right lines.

The theory could be used to find the pressure at which a solid "melts" at temperatures higher than the critical temperature, though such a change of phase would really be equivalent to sublimation. The theory we have developed for liquids could be applied to the dense gas in equilibrium with the solid at high temperatures. Whether or not there is a critical temperature above which a gas could not be compressed to the solid phase is a matter for further investigation. We hope to carry out numerical calculations to see what the present theory implies at high temperatures. It is at any rate clear that, if such a critical temperature does not exist, the theory will give an asymptotic formula for the temperature of melting at high pressures and may find useful applications in geophysics and other branches of physics.*

---

* The pressure at which melting occurs for a given temperature always lies between p_c' and $p_c' + p_c''$ where these pressures refer to the volume v_c defined in (3·46). Hence we could write $p_m = p_c' + \epsilon p_c''$ where ϵ lies between 0 and 1 and from this infer an asymptotic formula for the relation between melting temperature and pressure at high pressures.

The calculations of this paper together with those of Papers I and II give the free energy of an assembly as a function of volume and temperature through the whole of the solid, liquid and gaseous phases. If the free energy is plotted as a surface in terms of these variables, it is found that in the solid-liquid range it consists of two valleys separated by a ridge. Points on this ridge correspond to unstable states of the assembly, for a state of lower free energy can be attained by following the common tangent to the two valleys. Points on any such tangent correspond to a mixture of the two phases, solid and liquid. These common tangents develop a surface which is like a road, banked first in one direction (corresponding to melting at a negative pressure) and then in the other (melting at positive pressure). By means of such a surface it is possible to follow the changes consequent on an increase of temperature at constant pressure. At zero pressure, for example, the path of the representative point on the free energy surface follows at first the bottom of the valley which corresponds to the solid state, the volume increasing with increase of temperature. At the temperature of melting the point crosses the flat portion at the place where it is horizontal. This corresponds to melting and a discontinuous increase of volume. Thereafter it follows the bottom of the second valley. Thus we get successively thermal expansion of the solid, dilation on melting and finally thermal expansion of the liquid.

We are indebted to the Department of Scientific and Industrial Research for a grant to enable this research to be carried out.

APPENDIX

The boiling-points and heats of vaporization of gases

In Paper II we calculated the boiling-points of the inert gases from the atomic forces, assuming that the whole of the available volume was shared equally by all the atoms just as it would be in a perfect gas. According to the preceding paper this assumption is too drastic, for the extra term in the free energy A'' due to disorder does not reach a value as great as $-NkT$ owing to the increase of energy as well as entropy due to disorder (cf. fig. 3 above). We have therefore repeated the calculations of Paper II, replacing the free energy of the liquid A_l on the left-hand sides of equations (15) and (16) by $A'+A''$. The results are shown in Table IV where it will be seen that there is marked improvement in the agreement with experiment.

TABLE IV. CALCULATED BOILING-POINTS OF GASES

	Boiling temperature			Entropy of vaporization (cal./deg.)		
	Calc. Paper II	Calc. (new method)	Obs.	Calc. Paper II	Calc. new method)	Obs.
Argon	94·1	87·3	87·4	20·7	17·5	17·2
Nitrogen	79·0	75·0	77·2	19·8	16·7	17·3

REFERENCES

Bethe 1935 *Proc. Roy. Soc.* A, **150**, 552.
Bragg and Williams 1934 *Proc. Roy. Soc.* A, **145**, 699.
— — 1935 *Proc. Roy. Soc.* A, **151**, 540.
Bragg, Sykes and Bradley 1937 *Proc. Phys. Soc.* **49**, 96 (extra part).
Braunbek 1926 *Z. Phys.* **38**, 549.
Eyring 1936 *J. Chem. Phys.* **4**, 283.
Eyring and Hirschfelder 1937 *J. Phys. Chem.* **41**, 249.
Fowler 1936 "Statistical Mechanics", p. 804, Camb. Univ. Press.
Frenkel, J. 1937 *Trans. Faraday Soc.* **33**, 58.
Herzfeld and Göppert-Mayer 1934 *Phys. Rev.* **46**, 995.
Hirschfelder, Stevenson and Eyring 1937 *J. Chem. Phys.* **5**, 896.
Kincaid and Eyring 1937 *J. Chem. Phys.* **5**, 587.
Lennard-Jones 1937 *Physica*, (IV), **10**, 941.
Lennard-Jones and Devonshire 1937 *Proc. Roy. Soc.* A, **163**, 53, and **165**, 1.
Lennard-Jones and Ingham 1925 *Proc. Roy. Soc.* A, **107**, 636.
Lindemann 1910 *Phys. Z.* **11**, 609.
Newton and Eyring 1937 *Trans. Faraday Soc.* **33**, 73.
Peierls 1936 *Proc. Camb. Phil. Soc.* **32**, 471.
Raschevsky 1927 *Z. Phys.* **40**, 214.
Simon, Ruhemann and Edwards 1930 *Z. phys. Chem.* **6**, 331; **7**, 80.
Williams 1935 *Proc. Roy. Soc.* A, **152**, 231.

Collective electron ferromagnetism
II. Energy and specific heat

By Edmund C. Stoner, F.R.S.
Reader in Physics at the University of Leeds

(*Received* 24 *October* 1938)

1. Introduction

A treatment of collective electron ferromagnetism on the basis of Fermi-Dirac statistics was developed in a previous paper (Stoner 1938*a*, referred to as I). The problem considered was that of the magnetic characteristics associated with electrons in an energy band of standard form (in which the energy density of states is proportional to the square root of the energy), and subject to interchange interaction effects which involve dependence of the energy on the magnetization. The distribution of states is specified by a parameter ϵ_0, giving the maximum electron energy at absolute zero in the absence of the magnetization dependent part of the interchange interaction energy. The relative magnetization, ζ, the ratio of the number of parallel spins to the total number of potentially effective spins, varies with the temperature in a manner depending on the ratio $k\theta'/\epsilon_0$, where $k\theta'$ is a measure of the maximum quasi-magnetic interaction energy per electron. In the limit $\epsilon_0/k\theta' \to 0$, the expressions pass over into those obtained using classical statistics, and the Curie temperature, θ, at which the spontaneous magnetization becomes zero, is then equal to θ'. In general θ is less than θ'. Tables were given showing the dependence on ζ of kT/ϵ_0 (and also of T/θ) for a wide range of values of $k\theta'/\epsilon_0$.

In the discussion of the treatment of the magnetic characteristics, the desirability of a corresponding treatment of the energetic characteristics was pointed out. Such a treatment is given in the present paper. In a theoretical treatment of ferromagnetism consideration of the energy relations is particularly important, since the intrinsic magnetization, with which the basic theory is concerned, cannot usually be deduced, with any precision, from experimental determinations on the apparent magnetization alone. Complementary investigations are necessary on magneto-thermal effects, such as the enhanced specific heat of ferromagnetics both below and above the Curie point, and the magneto-caloric effect. A satisfactory theoretical treatment of ferromagnetism should lead to consistent conclusions when applied to the interconnected magnetic and energetic effects. The usual treatment, based on the use of classical statistics, is here unsatisfactory. The values deduced for the molecular field coefficient, for example, from the paramagnetism above the Curie point, from the excess specific heat below, and from the Curie temperature and the low-temperature saturation magnetization, differ widely from each other even when allowance is made for various possible secondary effects which make the apparent discrepancies still greater (Stoner 1936*a*). It is for reasons such as these that it seems desirable to explore in some detail the modifications introduced by the treatment of the problem on the basis of Fermi-Dirac statistics.

The treatment here given is necessarily idealized by the choice of a particular band form, but the choice has the advantage of making precise computations possible over the whole range of variation of relevant parameters, and of bringing out clearly the character of the modifications resulting from the changed statistical treatment, independently of any effects which might arise owing to special peculiarities in the band form.

From the general expressions, using tables of Fermi-Dirac functions (McDougall and Stoner 1938), numerical computations of the energy may be made, for a given value of $k\theta'/\epsilon_0$, at particular values of kT/ϵ_0 (or T/θ), provided that the corresponding values of ζ are known. For computations of the specific heat, which are desirable for purposes of comparison with experiment, the obvious procedure is to determine the energies at a series of equally spaced temperatures, from which the specific heat may be found by numerical differentiation. To obtain sets of values of the specific heat as a function of the temperature for each value of $k\theta'/\epsilon_0$ as complete as those for the magnetization would, however, involve an almost prohibitive amount of labour in computation. This is due to the fact that the previous results give values of T/θ at equal intervals of ζ, and these intervals are not small enough for the values of ζ at equal intervals of T/θ to be obtained with sufficient accuracy by the usual methods of inverse interpolation. In effect a number of extended calculations for trial interpolates have to be made for each value of T/θ. For the general purpose in view, however, it is doubtful whether the labour entailed in obtaining a complete set of specific heat values (that is at small intervals over the T/θ range) for each value of $k\theta'/\epsilon_0$ would be justified, partly because the number of ferromagnetics for which accurate specific heat data are available is very limited. Detailed computations of the specific heat at small intervals over the T/θ range have therefore been made only for one value of $k\theta'/\epsilon_0$, in addition to those for the classical limit, $\epsilon_0/k\theta' \rightarrow 0$. (For the limit $k\theta/\epsilon_0 \rightarrow 0$ the results can be expressed in simple analytical form and no detailed computations are required.) This complete set of results is supplemented by computations, for the whole range of values of $k\theta'/\epsilon_0$, of the main energy and specific heat characteristics—the energy changes between absolute zero and the Curie point, and the specific heat at low temperatures and just below and just above the Curie point. The detailed computations are made for the value $k\theta'/\epsilon_0 = 2^{-\frac{1}{3}}$ this value corresponding to an interchange interaction effect just sufficient to produce complete saturation at absolute zero. For smaller values of $k\theta'/\epsilon_0$ the dependence of ζ/ζ_0 on T/θ approaches more and more closely the relatively simple form which holds in the limit $k\theta'/\epsilon_0 \rightarrow \frac{2}{3}$ (corresponding to $k\theta/\epsilon_0 \rightarrow 0$), and the process of interpolation would become progressively less troublesome. For greater values of $k\theta'/\epsilon_0$, the classical form (which holds for $\epsilon_0/k\theta' \rightarrow 0$) is approached. The detailed computations are probably more troublesome for $k\theta'/\epsilon_0 = 2^{-\frac{1}{3}}$ than for any other value, but the results, in conjunction with those for the limiting values $k\theta/\epsilon_0 \rightarrow 0$ and $\epsilon_0/k\theta \rightarrow 0$, would be particularly useful in enabling approximate estimates to be made for any value of $k\theta'/\epsilon_0$ (or $k\theta/\epsilon_0$) over the whole range. Moreover, some preliminary considerations of the experimental results for iron, nickel, and cobalt (Stoner 1938*b*) indicate that for all these metals the effective interaction is in the neighbourhood of that required just to produce complete saturation at absolute zero, so that the choice is appropriate from the point of view of applications to actual metals.

After developing the theory as far as is necessary for the derivation of general expressions for the energy, a number of special theoretical expressions are obtained, in particular for the specific heat in the neighbourhood of the Curie point and at very low temperatures. The detailed computational results for $\epsilon_0/k\theta' \rightarrow 0$, and $k\theta'/\epsilon_0 = 2^{-\frac{1}{3}}$ are then presented, followed by results for the main energy and specific heat characteristics over the range of $k\theta'/\epsilon_0$ values. Finally, a general discussion is given. This is very brief, as it is hoped later to supplement these two papers by a discussion of the treatment as a whole, that is of magnetic as well as of energetic characteristics, in relation to experimental results, and to other theoretical treatments.

2. GENERAL THEORY AND METHOD OF COMPUTATION

The system under consideration is a set of N electrons in a partially filled energy band, and subject to interchange interaction effects which give rise to a term in the energy expression proportional to the square of the magnetization. The energy density of states is taken as proportional to the square root of the energy, apart from the magnetization dependent interaction energy. A method of dealing with the statistical problem has been developed in I, with particular reference to the derivation of general expressions for the magnetization. It will be sufficient here to indicate essential steps in the treatment, without detailed discussion. An appropriate starting point for the derivation of the energy expressions is the expression for the energy distribution of electrons, which may be put in the following form:

$$\frac{dN}{d\epsilon} = \frac{\nu(\epsilon)}{\exp\{(\epsilon+\epsilon')/kT-\eta\}+1}. \tag{2·1}$$

Here $\nu(\epsilon)$ is the energy density of states; with the distribution considered,

$$\nu(\epsilon) = \tfrac{3}{4}(N/\epsilon_0^{\frac{3}{2}})\,\epsilon^{\frac{1}{2}}, \tag{2·2}$$

where ϵ_0 is the maximum electron energy at absolute zero in the absence of the magnetization dependent part of the interchange interaction. The energy ϵ is taken to include the whole of the energy except that depending on the magnetization and on the external field. The corresponding terms make up ϵ'; with the previous symbolism,

$$\epsilon' = \pm\,(k\theta'\zeta+\mu H), \tag{2·3}$$

where ζ is the relative magnetization, that is

$$\zeta = M/N\mu, \tag{2·4}$$

where M is the resultant magnetic moment of the N electrons. Substituting (2·2) and (2·3) in (2·1), and using the abbreviations

$$x = \epsilon/kT; \quad \beta = k\theta'\zeta/kT = (\theta'/T)\zeta; \quad \beta' = \mu H/kT, \tag{2·5}$$

the distribution expression is obtained in the form

$$\frac{dN}{dx} = \tfrac{3}{4}N\left(\frac{kT}{\epsilon_0}\right)^{\frac{3}{2}}\left[\frac{x^{\frac{1}{2}}}{\exp\{x-(\eta+\beta+\beta')\}+1}+\frac{x^{\frac{1}{2}}}{\exp\{x-(\eta-\beta-\beta')\}+1}\right]. \tag{2·6}$$

Writing
$$F_k(\eta) = \int_0^\infty \frac{x^k\,dx}{e^{x-\eta}+1}, \tag{2·7}$$

integration of (2·6) gives

$$N = \tfrac{3}{4}N(kT/\epsilon_0)^{\frac{3}{2}}\,[F_{\frac{1}{2}}(\eta+\beta+\beta')+F_{\frac{1}{2}}(\eta-\beta-\beta')]. \tag{2·8}$$

The two terms in (2·8) give the numbers of electrons with spins parallel and antiparallel to the resultant field, and the expression for the magnetization follows immediately:

$$M = \tfrac{3}{4}N\mu(kT/\epsilon_0)^{\frac{3}{2}}\,[F_{\frac{1}{2}}(\eta+\beta+\beta')-F_{\frac{1}{2}}(\eta-\beta-\beta')]. \tag{2·9}$$

The equations (2·8) and (2·9) are the fundamental equations previously used, but they are here derived in a more direct manner. The relative magnetization is obtained by division:

$$\zeta = \frac{M}{N\mu} = \frac{F_{\frac{1}{2}}(\eta+\beta+\beta')-F_{\frac{1}{2}}(\eta-\beta-\beta')}{F_{\frac{1}{2}}(\eta+\beta+\beta')+F_{\frac{1}{2}}(\eta-\beta-\beta')}. \tag{2·10}$$

The total energy is not obtained simply by multiplying the first and second terms in (2·6) by $kT(x-\beta-\beta')$ and $kT(x+\beta+\beta')$ respectively, and integrating, for the energy corresponding to β is mutual energy of the electrons arising from interaction, and in this process it would be included twice. It is necessary, therefore, to subtract from the integral so obtained the interaction energy, $-\frac{1}{2}Nk\theta'\zeta^2$, or, equivalently, to use $\frac{1}{2}\beta$ in place of β in the multiplying factors. Making use of (2·8) and (2·10) the expression for the total energy, E, may be obtained in the form

$$\frac{E}{N\epsilon_0} = \left(\frac{kT}{\epsilon_0}\right)\left\{\frac{F_{\frac{3}{2}}(\eta+\beta+\beta')+F_{\frac{3}{2}}(\eta-\beta-\beta')}{F_{\frac{1}{2}}(\eta+\beta+\beta')+F_{\frac{1}{2}}(\eta-\beta-\beta')}-\beta'\zeta-\frac{1}{2}\left(\frac{\theta'}{T}\right)\zeta^2\right\}. \tag{2·11}$$

In the absence of an external field ($\beta'=0$), (2·11) reduces to

$$\frac{E}{N\epsilon_0} = \left(\frac{kT}{\epsilon_0}\right)\frac{F_{\frac{3}{2}}(\eta+\beta)+F_{\frac{3}{2}}(\eta-\beta)}{F_{\frac{1}{2}}(\eta+\beta)+F_{\frac{1}{2}}(\eta-\beta)}-\frac{1}{2}\left(\frac{k\theta'}{\epsilon_0}\right)\zeta^2, \tag{2·12}$$

this being the basic expression for the calculation of the energy below the Curie point. Above the Curie point, this reduces further to

$$\frac{E}{N\epsilon_0} = \left(\frac{kT}{\epsilon_0}\right)\frac{F_{\frac{3}{2}}(\eta)}{F_{\frac{1}{2}}(\eta)}, \tag{2·13}$$

the usual expression, which has already been applied in the computation of electronic energy and specific heat (see Stoner 1938*c*).

Although expressions for entropy and free energy are not required for the immediate purpose of this paper, it may be well to give them for completeness. They may be readily obtained by adapting the treatment of Nordheim (1934) to the present more complex problem. It is unnecessary to give details. The expression for the entropy, S, may be put in the form

$$\frac{S}{Nk} = \frac{3}{4}\left(\frac{kT}{\epsilon_0}\right)^{\frac{3}{2}}\left[\tfrac{5}{3}\{F_{\frac{3}{2}}(\eta+\beta+\beta')+F_{\frac{3}{2}}(\eta-\beta-\beta')\}\right.$$
$$\left.-(\eta+\beta+\beta')F_{\frac{1}{2}}(\eta+\beta+\beta')-(\eta-\beta-\beta')F_{\frac{1}{2}}(\eta-\beta-\beta')\right]. \tag{2·14}$$

The free energy, F, is then obtainable from the relation

$$F = E - TS.$$

After some transformations, the expression for F may be obtained in the form

$$\frac{F}{N\epsilon_0} = \left(\frac{kT}{\epsilon_0}\right)\eta - \frac{1}{2}\left(\frac{kT}{\epsilon_0}\right)^{\frac{5}{2}}[F_{\frac{3}{2}}(\eta+\beta+\beta')+F_{\frac{3}{2}}(\eta-\beta-\beta')]+\frac{1}{2}\left(\frac{k\theta'}{\epsilon_0}\right)\zeta^2. \qquad (2{\cdot}15)$$

In the expression given in I (2·2 and 2·8) the last term was incorrectly omitted. This, however, did not affect the application being made, in which the magnetization was derived from the relation

$$M = -\partial F/\partial H.$$

It may be noted that, by using the complete expression, the magnetization in zero field ($\beta' = 0$) may be derived directly by applying the relation

$$\partial F/\partial\zeta = 0.$$

General method of computation. The computations of energy and specific heat will be limited to the case in which the external field is zero. It is convenient to regard the energy as made up of two parts corresponding to the two terms in (2·12) and to write

$$E = E_E + E_M, \qquad (2{\cdot}16)$$

where $\quad E_E = NkT\{F_{\frac{3}{2}}(\eta+\beta)+F_{\frac{3}{2}}(\eta-\beta)\}/\{F_{\frac{1}{2}}(\eta+\beta)+F_{\frac{1}{2}}(\eta-\beta)\},$

$$E_M = -\tfrac{1}{2}Nk\theta'\zeta^2.$$

The part E_M is the "molecular field" energy, corresponding to the quasi-magnetic interaction. The part E_E is the energy corresponding to the electronic distribution among the available states. With classical statistics, E_E is independent of ζ, but with Fermi-Dirac statistics this does not hold (except in the classical limit). The specific heat* will be denoted by C, the subdivision into two parts again being convenient:

$$C = \frac{dE}{dT} = C_E + C_M = \frac{dE_E}{dT} + \frac{dE_M}{dT}. \qquad (2{\cdot}17)$$

It is appropriate to consider a reduced energy, $E/N\epsilon_0$, and a reduced specific heat, C/Nk, the relation between them being

$$C/Nk = d(E/N\epsilon_0)/d(kT/\epsilon_0). \qquad (2{\cdot}18)$$

---

* Strictly, C does not denote a "specific" heat, but a heat capacity, the contribution to the heat capacity of the material from the N electrons under consideration. The term "specific" heat is used here in a general sense in the interests of brevity, and should give rise to no misunderstanding, since the relation of C or C/Nk to specific heat as ordinarily understood is obvious.

In a complete computation of the energy as a function of the temperature for a particular value of $k\theta'/\epsilon_0$, it is appropriate to calculate the energy for suitably spaced values of T/θ (say $T/\theta = 0{\cdot}00$, $0{\cdot}05$, $0{\cdot}10$, ...). The previous tables give the values of kT/ϵ_0, and of T/θ at equal intervals of ζ ($\zeta = 0{\cdot}0$, $0{\cdot}1$, $0{\cdot}2$, ...). If the value of ζ is known for a particular value of T/θ, the corresponding values of $F_{\frac{1}{2}}(\eta+\beta)$ and $F_{\frac{1}{2}}(\eta-\beta)$ are found from the equations (I, 4·7 and 4·8)

$$F_{\frac{1}{2}}(\eta \pm \beta) = \frac{2}{3}\left(\frac{\theta}{T}\right)^{\frac{3}{2}}\left(\frac{\epsilon_0}{k\theta}\right)^{\frac{3}{2}}(1 \pm \zeta). \tag{2·19}$$

Inverse interpolation in the $F_{\frac{1}{2}}(\eta)$ table gives $\eta+\beta$ and $\eta-\beta$, and direct interpolation in the $F_{\frac{3}{2}}(\eta)$ table gives $F_{\frac{3}{2}}(\eta+\beta)$ and $F_{\frac{3}{2}}(\eta-\beta)$ required in (2·12).

Over most of the T/θ range it is not possible to obtain sufficiently accurate values of ζ for a particular T/θ from the previous tables by the usual inverse interpolation methods, or by the application of approximate series representations. It is then necessary to use a method of trial. An estimated ζ for the particular $k\theta'/\epsilon_0$ and T/θ is used in (2·19), and the corresponding β is found; from this value of β, $k\theta'/\epsilon_0$ is calculated, using (2·5). Revised estimates of ζ are then made until there is agreement with the initial $k\theta'/\epsilon_0$ to the requisite degree of accuracy. It is this preliminary determination of ζ values which makes so lengthy the computation of a series of energy values sufficiently complete and accurate for the subsequent calculation of specific heat by numerical differentiation.

For $T/\theta \geqslant 1$, the energy may be calculated using (2·13) together with (2·19), which simplifies to

$$F_{\frac{1}{2}}(\eta) = \frac{2}{3}\left(\frac{\theta}{T}\right)^{\frac{3}{2}}\left(\frac{\epsilon_0}{k\theta}\right)^{\frac{3}{2}}. \tag{2·20}$$

Alternatively, and more conveniently, the values of the energy may be obtained by interpolation in the tables already published (Stoner 1938*c*) giving $E/N\epsilon_0$ as a function of kT/ϵ_0. Further tables give the reduced specific heat, C/Nk, as a function of kT/ϵ_0.

3. The classical and Fermi-Dirac limits

(i) *Classical statistics.* In the limit $\epsilon_0/k\theta \to 0$ the results given by the present treatment pass over into those obtained using classical statistics. The method of determining ζ as a function of T/θ is described in I (see equations 4·4 to 4·6). In the limit considered, $-\eta \gg 1$, and

$$F_{\frac{1}{2}}(\eta) \to \tfrac{1}{2}\pi^{\frac{1}{2}}e^{\eta}, \quad F_{\frac{3}{2}}(\eta) \to \tfrac{3}{4}\pi^{\frac{1}{2}}e^{\eta}.$$

Further (see I, 5·2)

$$\theta' \to \theta.$$

The expression (2·12) for the energy becomes

$$E/N = \tfrac{3}{2}kT - \tfrac{1}{2}k\theta\zeta^2, \tag{3·1}$$

giving, for the specific heat,

$$C/Nk = \tfrac{3}{2} - \tfrac{1}{2}d\zeta^2/d(T/\theta). \tag{3·2}$$

The second term corresponds to the "magnetic" specific heat:

$$C_M/Nk = -\tfrac{1}{2}d\zeta^2/d(T/\theta). \tag{3·3}$$

The constant term, corresponding to C_E, has usually been ignored in classical treatments, either tacitly, or with the doubtful "empirical" justification on which classical treatments had frequently to rely.

The magnetization dependent part of the energy, E_M, is conveniently considered in the reduced form,

$$E_M/Nk\theta = -\tfrac{1}{2}\zeta^2, \tag{3·4}$$

and it is obtainable at once from the previous results (I, Table I) giving ζ as a function of T/θ. From a set of values of $E_M/Nk\theta$ at equal intervals of T/θ, values of C_M/Nk may be found by numerical differentiation. A direct expression for C_M/Nk in terms of ζ and T/θ is, however, more convenient. Using the relation (I, 4·6)

$$\frac{\theta}{T} = \frac{1}{2\zeta}\log\frac{1+\zeta}{1-\zeta},$$

a direct expression is easily obtained:

$$\frac{C_M}{Nk} = -\frac{1}{2}\frac{d\zeta^2}{d(T/\theta)} = -\frac{\zeta}{d(T/\theta)/d\zeta}$$

$$= \frac{\zeta^2(1-\zeta^2)}{(T/\theta)\{(T/\theta)-(1-\zeta^2)\}}. \tag{3·5}$$

This formula is particularly advantageous for the initial part of the T/θ range, where it is difficult to obtain accurate values by the numerical differentiation method.

(ii) *Fermi-Dirac limit* ($k\theta/\epsilon_0 \to 0$). It has been shown in I that spontaneous magnetization occurs only for $k\theta'/\epsilon_0 > \tfrac{2}{3}$, and that for $k\theta'/\epsilon_0 < 2^{-\frac{1}{3}}$ ($= 0·793\,701$) the relative magnetization at absolute zero, ζ_0, is less than unity. For $k\theta/\epsilon_0 \to 0$, the following relation is obtained, retaining terms in the series expansion up to those in $(kT/\epsilon_0)^4$ and ζ^4 (cf. I, 5·8):

$$\frac{k\theta'}{\epsilon_0} = \frac{2}{3}\left[1 + \frac{2}{27}\zeta^2\left(1+\frac{7}{18}\zeta^2\right) + \frac{\pi^2}{12}\left(\frac{kT}{\epsilon_0}\right)^2\left(1+\frac{20}{27}\zeta^2\right) + \frac{3\pi^4}{80}\left(\frac{kT}{\epsilon_0}\right)^4\right]. \tag{3·6}$$

This relation includes that between $k\theta'/\epsilon_0$ and $k\theta/\epsilon_0$ (cf. I, 5·3):

$$\frac{k\theta'}{\epsilon_0} = \frac{2}{3}\left[1 + \frac{\pi^2}{12}\left(\frac{k\theta}{\epsilon_0}\right)^2 + \frac{3\pi^4}{80}\left(\frac{k\theta}{\epsilon_0}\right)^4\right]. \tag{3·7}$$

These two equations lead to an explicit expression for ζ^2 in terms of $k\theta/\epsilon_0$ and kT/ϵ_0,

$$\zeta^2 = \frac{27}{2}\frac{\pi^2}{12}\left\{\left(\frac{k\theta}{\epsilon_0}\right)^2 - \left(\frac{kT}{\epsilon_0}\right)^2\right\}\left\{1 + \frac{\pi^2}{80}\left(\frac{k\theta}{\epsilon_0}\right)^2 + \frac{13\pi^2}{240}\left(\frac{kT}{\epsilon_0}\right)^2\right\}, \tag{3·8}$$

which includes the relation between ζ_0 and $k\theta/\epsilon_0$,

$$\zeta_0^2 = \frac{27}{2}\frac{\pi^2}{12}\left(\frac{k\theta}{\epsilon_0}\right)^2\left\{1 + \frac{\pi^2}{80}\left(\frac{k\theta}{\epsilon_0}\right)^2\right\}. \tag{3·9}$$

In the limit considered, $\eta \pm \beta$ may be obtained in terms of kT/ϵ_0 and ζ from (2·19), using the inverse asymptotic series, and substituted in the asymptotic series for $F_{\frac{3}{2}}(\eta \pm \beta)$ in (2·16). For the distributional part of the energy, the following expression is obtained, to the same approximation as above:

$$\frac{E_E}{N\epsilon_0} = \frac{3}{5}\left[1 + \frac{5}{9}\zeta^2\left(1 + \frac{1}{27}\zeta^2\right) + \frac{5\pi^2}{12}\left(\frac{kT}{\epsilon_0}\right)^2\left(1 - \frac{1}{9}\zeta^2\right) - \frac{\pi^4}{16}\left(\frac{kT}{\epsilon_0}\right)^4\right]. \tag{3·10}$$

Substitution of (3·7) and (3·8) in the general expression (cf. 2·16) for the reduced quasi-magnetic part of the energy,

$$\frac{E_M}{N\epsilon_0} = -\frac{1}{2}\left(\frac{k\theta'}{\epsilon_0}\right)\zeta^2, \tag{3·11}$$

gives, as an appropriate expression in terms of $k\theta/\epsilon_0$ and kT/ϵ_0,

$$\frac{E_M}{N\epsilon_0} = -\frac{3\pi^2}{8}\left\{\left(\frac{k\theta}{\epsilon_0}\right)^2 - \left(\frac{kT}{\epsilon_0}\right)^2\right\}\left\{1 + \frac{23\pi^2}{240}\left(\frac{k\theta}{\epsilon_0}\right)^2 + \frac{13\pi^2}{240}\left(\frac{kT}{\epsilon_0}\right)^2\right\}. \tag{3·12}$$

The magnetic specific heat is given by

$$\frac{C_M}{Nk} = \frac{d(E_M/N\epsilon_0)}{d(kT/\epsilon_0)} = \frac{3\pi^2}{4}\left(\frac{kT}{\epsilon_0}\right)\left\{1 + \frac{\pi^2}{24}\left(\frac{k\theta}{\epsilon_0}\right)^2 + \frac{13\pi^2}{120}\left(\frac{kT}{\epsilon_0}\right)^2\right\}, \tag{3·13}$$

the limiting value being

$$\frac{C_M}{Nk} = \frac{3\pi^2}{4}\left(\frac{kT}{\epsilon_0}\right) = \frac{2^{\frac{1}{2}}\pi\zeta_0}{2}\left(\frac{T}{\theta}\right) = 2{\cdot}221\,441\zeta_0(T/\theta). \tag{3·14}$$

It is convenient to treat the "distributional" specific heat, C_E, as a sum of two parts, C_1 and C_2, say, the first corresponding to the change of temperature with ζ constant, and the second to the change of ζ with temperature:

$$\frac{C_E}{Nk} = \frac{d(E_E/N\epsilon_0)}{d(kT/\epsilon_0)} = \frac{C_1}{Nk} + \frac{C_2}{Nk} = \left\{\frac{\partial(E_E/N\epsilon_0)}{\partial(kT/\epsilon_0)}\right\}_\zeta + \left\{\frac{\partial(E_E/N\epsilon_0)}{\partial\zeta}\right\}_{kT/\epsilon_0} \times \frac{d\zeta}{d(kT/\epsilon_0)}. \qquad (3\cdot15)$$

The following expressions are obtained from (3·10), making use of (3·8):

$$\frac{C_1}{Nk} = \frac{\pi^2}{2}\left(\frac{kT}{\epsilon_0}\right)\left\{1 - \frac{\pi^2}{8}\left(\frac{k\theta}{\epsilon_0}\right)^2 - \frac{7\pi^2}{40}\left(\frac{kT}{\epsilon_0}\right)^2\right\}, \qquad (3\cdot16)$$

$$\frac{C_2}{Nk} = -\frac{3\pi^2}{4}\left(\frac{kT}{\epsilon_0}\right)\left\{1 + \frac{\pi^2}{24}\left(\frac{k\theta}{\epsilon_0}\right)^2 - \frac{7\pi^2}{120}\left(\frac{kT}{\epsilon_0}\right)^2\right\}. \qquad (3\cdot17)$$

Substitution for ζ^2 from (3·8) is legitimate only for $T < \theta$. Above the Curie point, C_M and C_2 are zero, as is clear from the basic form of the expressions, and C_1 reduces to

$$\left(\frac{C_1}{Nk}\right)_{T/\theta>1} = \frac{\pi^2}{2}\left(\frac{kT}{\epsilon_0}\right) - \frac{3\pi^4}{20}\left(\frac{kT}{\epsilon_0}\right)^3, \qquad (3\cdot18)$$

the usual low temperature specific heat expression. In the limit $k\theta/\epsilon_0 \to 0$,

$$C_2 \to -C_M \quad \text{and} \quad C \to C_1.$$

The compensation of the magnetic specific heat by the part C_2 of C_E holds generally for $T/\theta \to 0$, but with increasing T/θ there is a residual excess. From (3·13) and (3·17),

$$\frac{C_M + C_2}{Nk} = \frac{\pi^4}{8}\left(\frac{kT}{\epsilon_0}\right)^3, \qquad (3\cdot19)$$

this relation holding up to the Curie point. The value at the Curie point gives the limiting value of the specific heat discontinuity:

$$\left\{\left(\frac{\Delta C}{Nk}\right)_{T/\theta=1}\right\}_{k\theta/\epsilon_0\to0} = \frac{\pi^4}{8}\left(\frac{k\theta}{\epsilon_0}\right)^3 = 12\cdot176\,136\left(\frac{k\theta}{\epsilon_0}\right)^3. \qquad (3\cdot20)$$

4. THE CURIE POINT REGION AND THE LOW-TEMPERATURE LIMIT

(i) *The Curie point region.* The rise in the specific heat to a maximum followed by an abrupt decrease in the Curie point region is one of the salient experimental characteristics of ferromagnetics. The magnitude of the specific heat discontinuity for the sequence of idealized ferromagnetics under consideration could be found by detailed computations of the energy in the neighbourhood of the Curie point; in view of the labour that this would entail, however, expressions from which the discontinuity can be more directly computed are clearly desirable. Such expressions may be obtained by making use of Taylor series expansions, so that, in particular, the specific heat immediately below and above the Curie point may be expressed in terms of $F_{\frac{3}{2}}(\eta)$ and $F_{\frac{1}{2}}(\eta)$ and its derivatives at the value of η corresponding to $T=\theta$. It will be sufficient to give the main steps in the derivation of the formulae.

Let $F_{\frac{3}{2}}$, F, F', ... refer to the values of the functions $F_{\frac{3}{2}}(\eta)$, $F_{\frac{1}{2}}(\eta)$, $dF_{\frac{1}{2}}(\eta)/d\eta$, ... for $\eta=\eta_\theta$, i.e. corresponding to $T=\theta$. For $\beta'=0$, (2·8) may be written

$$F_{\frac{1}{2}}(\eta+\beta)+F_{\frac{1}{2}}(\eta-\beta) = \tfrac{4}{3}(\epsilon_0/kT)^{\frac{3}{2}}, \tag{4·1}$$

giving

$$F = \tfrac{2}{3}(\epsilon_0/k\theta)^{\frac{3}{2}}.$$

Writing

$$\eta = \eta_\theta+\alpha,$$

$$y = 1-(T/\theta),$$

a Taylor series expansion of (4·1) gives, to the required degree of approximation, for $y\to 0$, $\alpha/\eta_\theta\to 0$, $\beta\to 0$,

$$\begin{aligned}\alpha F'+\tfrac{1}{2}(\alpha^2+\beta^2)\,F'' &= \tfrac{2}{3}\{(\epsilon_0/kT)^{\frac{3}{2}}-(\epsilon_0/k\theta)^{\frac{3}{2}}\}\\ &= \{(\theta/T)^{\frac{3}{2}}-1\}\,F = \tfrac{3}{2}yF,\end{aligned}$$

so that

$$\alpha = \tfrac{3}{2}(F/F')\,y-\tfrac{1}{2}(F''/F')\,\beta^2.$$

Substitution of this expression for α in a series expansion of (2·10) (with $\beta'=0$) gives

$$\zeta = \beta\frac{F'}{F}\left[1+\frac{3}{2}\frac{F}{F'}\left\{\frac{F''}{F'}-\frac{F'}{F}\right\}y+\left\{\frac{1}{6}\frac{F'''}{F'}-\frac{1}{2}\left(\frac{F''}{F'}\right)^2\right\}\beta^2\right].$$

Since $\zeta/\beta = (T/\theta') = (\theta/\theta')(1-y) = (F'/F)(1-y)$,

$$\frac{\beta^2}{y} = \frac{9FF''-3(F')^2}{3(F'')^2-F'F'''}, \tag{4·2}$$

and

$$\zeta^2 = (F'/F)^2\beta^2.$$

An expression is thus obtained for the variation of magnetization with temperature near the Curie point, which may be put in the form

$$-\left[\frac{d\zeta^2}{d(T/\theta)}\right]_{T/\theta\to 1} = \left(\frac{F'}{F}\right)^2\frac{9FF''-3(F')^2}{3(F'')^2-F'F'''}. \tag{4·3}$$

This general expression includes as special cases the formulae previously given (I, 5·15 and 5·16) for $\epsilon_0/k\theta\to 0$ and $k\theta/\epsilon_0\to 0$, which may be readily obtained by utilizing the series expansions appropriate for the Fermi-Dirac functions in the two limits.

A Taylor series expansion of (2·12) gives for the energy in the neighbourhood of the Curie point the expression

$$\frac{E}{Nk\theta}=\frac{F_{\frac{3}{2}}}{F}+\left\{\frac{9}{4}\frac{F}{F'}-\frac{5}{2}\frac{F_{\frac{3}{2}}}{F}\right\}y+\frac{3}{4}\left\{\frac{F'}{F}-\frac{F''}{F'}\right\}\left(\frac{F}{F'}\right)^2\zeta^2-\frac{1}{2}\left(\frac{F}{F'}\right)\zeta^2, \qquad (4{\cdot}4)$$

in which the first term gives the energy at the Curie point. As before, the specific heat, C, may be treated as the sum of C_E and C_M, and C_E divided into two parts C_1 and C_2:

$$C=C_E+C_M=C_1+C_2+C_M=Nk\frac{d(E/Nk\theta)}{d(T/\theta)}=-Nk\frac{d(E/Nk\theta)}{dy}. \qquad (4{\cdot}5)$$

The partial contributions C_1, C_2 and C_M, correspond to the second, third and fourth terms in (4·4), and are given by the following expressions:

$$\left.\begin{aligned}\frac{C_1}{Nk}&=\frac{5}{2}\frac{F_{\frac{3}{2}}}{F}-\frac{9}{4}\frac{F}{F'},\\ \frac{C_2}{Nk}&=-\frac{3}{4}\left\{\frac{F'}{F}-\frac{F''}{F'}\right\}\left\{\frac{9FF''-3(F')^2}{3(F'')^2-F'F'''}\right\},\\ \frac{C_M}{Nk}&=\frac{1}{2}\left(\frac{F'}{F}\right)\left\{\frac{9FF''-3(F')^2}{3(F'')^2-F'F'''}\right\}.\end{aligned}\right\} \qquad (4{\cdot}6)$$

The part C_1 alone, the usual specific heat expression, gives the specific heat immediately above the Curie point (at $\theta+0$), and the sum of the three parts the specific heat immediately below (at $\theta-0$). For the discontinuity,

$$\frac{\Delta C}{Nk}=\frac{C_2+C_M}{Nk}=\left\{\frac{3}{4}\frac{F''}{F'}-\frac{1}{4}\frac{F'}{F}\right\}\left\{\frac{9FF''-3(F')^2}{3(F'')^2-F'F'''}\right\}. \qquad (4{\cdot}7)$$

The forms taken by these expressions as the classical and Fermi-Dirac limits are approached are of some interest. They may be obtained from the usual series expansions of the functions. For $(\epsilon_0/k\theta)\rightarrow 0$, using the abbreviation

$$z=\left(\frac{2}{\pi}\right)^{\frac{1}{2}}\left(\frac{\epsilon_0}{k\theta}\right)^{\frac{3}{2}}=0{\cdot}797\,885\left(\frac{\epsilon_0}{k\theta}\right)^{\frac{3}{2}},$$

and expanding to the term in z, the following expressions are obtained:

$$\left.\begin{aligned}C_1/Nk&=3/2-z/8,\\ C_2/Nk&=-3z/4,\\ C_M/Nk&=3/2+z/4,\\ \Delta C/Nk&=3/2-z/2.\end{aligned}\right\} \qquad (4{\cdot}8)$$

For $(k\theta/\epsilon_0)\to 0$, the series representations involve odd powers of $(k\theta/\epsilon_0)$. They are given below to the term in $(k\theta/\epsilon_0)^3$:

$$\left.\begin{aligned}
\frac{C_1}{Nk} &= \frac{\pi^2}{2}\left(\frac{k\theta}{\epsilon_0}\right)\left\{1-\frac{3\pi^2}{10}\left(\frac{k\theta}{\epsilon_0}\right)^2\right\},\\
\frac{C_2}{Nk} &= -\frac{3\pi^2}{4}\left(\frac{k\theta}{\epsilon_0}\right)\left\{1-\frac{\pi^2}{60}\left(\frac{k\theta}{\epsilon_0}\right)^2\right\},\\
\frac{C_M}{Nk} &= \frac{3\pi^2}{4}\left(\frac{k\theta}{\epsilon_0}\right)\left\{1+\frac{3\pi^2}{20}\left(\frac{k\theta}{\epsilon_0}\right)^2\right\},\\
\frac{\Delta C}{Nk} &= \frac{\pi^4}{8}\left(\frac{k\theta}{\epsilon_0}\right)^3.
\end{aligned}\right\} \tag{4·9}$$

These expressions are in agreement with the more general expressions obtained by a different method in § 3 (equations 3·16, 3·17, 3·13 and 3·20). For computational purposes it may be noted that

$$\frac{\pi^2}{2} = 4{\cdot}934\,802.$$

The remarkable modification in the specific heat relations resulting from the replacement of classical by Fermi-Dirac statistics is made particularly clear by these limiting forms for the specific heat at the Curie point. It is noteworthy that the discontinuity does not, in general, depend on the "magnetic" specific heat alone, so that it does not give an immediate measure of the interchange interaction effect (molecular field coefficient) as has previously been supposed in making application of a classical treatment (Stoner 1936*a*, *b*).

(ii) *The low-temperature limit.* Expressions for the energy at absolute zero are required primarily with a view to obtaining values for the total energy change up to the Curie point. The limiting forms for the specific heat are also desirable, since it is experimentally possible to determine the purely electronic specific heat at very low temperatures with considerable accuracy.

The relations which hold when the relative magnetization at absolute zero is less than unity $(\zeta_0<1)$ will first be considered. For $kT/\epsilon_0\to 0$, from (2·19), using the inverse asymptotic series (I, 2·16),

$$\eta\pm\beta = \left(\frac{\epsilon_0}{kT}\right)(1\pm\zeta)^{\frac{2}{3}}\left\{1-\frac{\pi^2}{12}\left(\frac{kT}{\epsilon_0}\right)^2(1\pm\zeta)^{-\frac{4}{3}}\right\}. \tag{4·10}$$

Substitution in the asymptotic series expansions for $F_{\frac{3}{2}}(\eta \pm \beta)$ in (2·16) gives

$$\frac{E_E}{N\epsilon_0} = \frac{3}{10}\left[(1+\zeta)^{\frac{5}{3}}\left\{1+\frac{5\pi^2}{12}\left(\frac{kT}{\epsilon_0}\right)^2(1+\zeta)^{-\frac{4}{3}}\right\}\right.$$
$$\left. +(1-\zeta)^{\frac{5}{3}}\left\{1+\frac{5\pi^2}{12}\left(\frac{kT}{\epsilon_0}\right)^2(1-\zeta)^{-\frac{4}{3}}\right\}\right]. \quad (4·11)$$

Thus, for the energy at absolute zero,

$$\frac{E}{N\epsilon_0} = \frac{E_E}{N\epsilon_0} + \frac{E_M}{N\epsilon_0} = \frac{3}{10}\{(1+\zeta_0)^{\frac{5}{3}} + (1-\zeta_0)^{\frac{5}{3}}\} - \frac{1}{2}\left(\frac{k\theta'}{\epsilon_0}\right)\zeta_0^2. \quad (4·12)$$

This expression is valid both for $\zeta_0 < 1$ $(k\theta'/\epsilon_0 < 2^{-\frac{1}{3}})$ and $\zeta_0 = 1$ $(k\theta'/\epsilon_0 \geqslant 2^{-\frac{1}{3}})$, the expression for E_E in the second range reducing to

$$\left(\frac{E_E}{N\epsilon_0}\right)_{\zeta_0=1} = \tfrac{3}{10} \times 2^{\frac{5}{3}} = 0·952\,440\,7. \quad (4·13)$$

The following expressions are obtained for the contributions to the specific heat:

$$\left.\begin{aligned}
\frac{C_1}{Nk} &= \frac{\pi^2}{4}\left(\frac{kT}{\epsilon_0}\right)\{(1+\zeta)^{\frac{1}{3}} + (1-\zeta)^{\frac{1}{3}}\},\\
\frac{C_2}{Nk} &= \tfrac{1}{2}\{(1+\zeta)^{\frac{2}{3}} - (1-\zeta)^{\frac{2}{3}}\}\left\{1-\frac{\pi^2}{24}\left(\frac{kT}{\epsilon_0}\right)^2(1-\zeta^2)^{-\frac{2}{3}}\right\}\frac{d\zeta}{d(kT/\epsilon_0)},\\
\frac{C_M}{Nk} &= -\frac{k\theta'}{\epsilon_0}\zeta\frac{d\zeta}{d(kT/\epsilon_0)}.
\end{aligned}\right\} \quad (4·14)$$

In the limit, the expression for C_1 gives

$$\left[\left(\frac{C_1}{Nk}\right)\bigg/\left(\frac{kT}{\epsilon_0}\right)\right]_{kT/\epsilon_0\to 0} = \left[\left(\frac{C_1}{Nk}\right)\bigg/\left(\frac{k\theta}{\epsilon_0}\right)\left(\frac{T}{\theta}\right)\right]_{T/\theta\to 0} = \frac{\pi^2}{4}\{(1+\zeta_0)^{\frac{1}{3}} + (1-\zeta_0)^{\frac{1}{3}}\}, \quad (4·15)$$

which is valid for $\zeta_0 = 1$, then reducing to

$$\left[\left(\frac{C_1}{Nk}\right)\bigg/\left(\frac{kT}{\epsilon_0}\right)\right]_{kT/\epsilon_0\to 0} = \frac{2^{\frac{1}{3}}\pi^2}{4} = 3·108\,731. \quad (4·16)$$

A comparison with (I, 5·9) shows that, for $\zeta_0 < 1$, C_2/Nk may be expressed as

$$\frac{C_2}{Nk} = +\frac{k\theta'}{\epsilon_0}\zeta\frac{d\zeta}{d(kT/\epsilon_0)}\left\{1-\frac{\pi^2}{8}\left(\frac{kT}{\epsilon_0}\right)^2(1-\zeta^2)^{-\frac{2}{3}}\right\}, \quad (4·17)$$

so that, for $kT/\epsilon_0 \to 0$, $\quad C_2 \to -C_M. \quad$ (4·18)

The limiting form for the total specific heat, C, is therefore the same as that for C_1; that is, the total specific heat near absolute zero is correctly given by the usual expression corresponding to the distribution, and in the limit there is no contribution to the specific heat arising from the change in the relative numbers of oppositely directed spins with temperature.

Expressions analogous to (4·11) and (4·14) may also be obtained for $\zeta_0 = 1$ $(k\theta'/\epsilon_0 > 2^{-\frac{1}{3}})$ by utilizing the series appropriate for $\eta - \beta < 0$. For $\epsilon_0/kT \gg 1$, $(\epsilon_0/kT)^{\frac{3}{2}}(1-\zeta) \to 0$,

$$\frac{E_K}{N\epsilon_0} = \tfrac{3}{10}(1+\zeta)^{\frac{5}{3}}\left\{1 + \frac{5\pi^2}{12}\left(\frac{kT}{\epsilon_0}\right)^2 (1+\zeta)^{-\frac{4}{3}}\right\} + \frac{3}{4}\left(\frac{kT}{\epsilon_0}\right)(1-\zeta), \tag{4·19}$$

$$\left.\begin{aligned} \frac{C_1}{Nk} &= \frac{\pi^2}{4}\left(\frac{kT}{\epsilon_0}\right)(1+\zeta)^{\frac{1}{3}} + \tfrac{3}{4}(1-\zeta), \\ \frac{C_2}{Nk} &= \left[\tfrac{1}{2}(1+\zeta)^{\frac{2}{3}}\left\{1 + \frac{\pi^2}{12}\left(\frac{kT}{\epsilon_0}\right)^2 (1+\zeta)^{-\frac{4}{3}}\right\} - \frac{3}{4}\left(\frac{kT}{\epsilon_0}\right)\right]\frac{d\zeta}{d(kT/\epsilon_0)}, \\ \frac{C_M}{Nk} &= -\frac{k\theta'}{\epsilon_0}\zeta\frac{d\zeta}{d(kT/\epsilon_0)}. \end{aligned}\right\} \tag{4·20}$$

The limiting values of $E_K/N\epsilon_0$ and of C_1/Nk for $\zeta \to \zeta_0 = 1$ are those given above. In the limit, for $k\theta'/\epsilon_0 > 2^{-\frac{1}{3}}$, C_2 and C_M vanish to a higher order than (kT/ϵ_0), since, as shown in I (5·21), under these conditions

$$\zeta = 1 - \frac{3}{2}\left(\frac{kT}{\epsilon_0}\right)^{\frac{3}{2}} \frac{\pi^{\frac{1}{2}}}{2} \exp\left\{-\frac{2\epsilon_0}{kT}\left(\frac{k\theta'}{\epsilon_0} - 2^{-\frac{1}{3}}\right)\right\}.$$

Thus for $kT/\epsilon_0 \to 0$, $\qquad C \to C_1.$ (4·21)

For the particular value $k\theta'/\epsilon_0 = 2^{-\frac{1}{3}}$, the sum $C_2 + C_M$ vanishes to the order kT/ϵ_0, so that for all values of $k\theta'/\epsilon_0$, the following relation holds:

$$\left[\left(\frac{C}{Nk}\right)\Big/\left(\frac{kT}{\epsilon_0}\right)\right]_{kT/\epsilon_0 \to 0} \to \frac{\pi^2}{4}\{(1+\zeta_0)^{\frac{1}{3}} + (1-\zeta_0)^{\frac{1}{3}}\}. \tag{4·22}$$

The splitting up of C_E into the two parts C_1 and C_2 is convenient in discussing the relations in the low temperature limit, but in the general method of computation it is C_E which is determined. The special method of calculating the limiting form of C is very useful, as the accurate calculation of C_E and C_M at low temperatures by the numerical differentiation method is particularly troublesome.

5. TEMPERATURE VARIATION OF ENERGY AND SPECIFIC HEAT FOR PARTICULAR VALUES OF $k\theta'/\epsilon_0$

(i) *Classical statistics*: $\epsilon_0/k\theta \to 0$. A table has already been given (Table I of I) of the magnetization, ζ, as a function of T/θ in the classical limit. For a complete set of values at intervals of 0·05 in T/θ, some additional computations were required. When ζ is known, the quasi-magnetic energy is readily obtained from (3·4). Differencing of the table (which is much more satisfactory as a check than differencing of ζ) revealed an error in the previous value of ζ for $T/\theta = 0{\cdot}60$. (The sixth decimal place digit was given as 7 instead of 2.) All the values of ζ were therefore recomputed to the degree of accuracy permitted by the Hayashi (1926) 8-decimal place tables for $\tanh x$. No other errors were found, though the additional figures resulted in an increase in one unit in the last place in the rounded six place values of ζ for $T/\theta = 0{\cdot}90$ and 0·95. The revised values of ζ are given in Table I, with additional values to complete the table at 0·05 intervals. For $T/\theta \leqslant 0{\cdot}20$, values of ζ are more accurately and conveniently calculated from the expression, valid for $T/\theta \to 0$,

$$1 - \zeta = 2e^{-2\theta/T},$$

than by use of the Hayashi tables.

Values of the specific heat were calculated from the direct expression (3·5) and also by numerical differentiation. It is the directly calculated (rounded) values which are given in Table I. Numerical differentiation, using seven-place values of ζ^2 and differences up to the fifth, gives values which are in agreement to within 2 units in the sixth place for $T/\theta \geqslant 0{\cdot}45$; but for smaller values of T/θ, the higher order differences become large, and the first differences small, and the lack of precision increases rapidly. The directly calculated values should be correct to 1 unit in the sixth place as given, though a satisfactory check by the method of differences cannot be applied in the first half of the table.

It is the quasi-magnetic parts only of the energy and specific heat which are given in the table. From (3·1) and (3·2), the values of the electronic distribution parts correspond to $E_E/Nk\theta = \frac{3}{2}(T/\theta)$, and $C_E/Nk = \frac{3}{2}$, these values holding also for the total electronic energy and specific heat above the Curie point.

(ii) $k\theta'/\epsilon_0 = 2^{-\frac{1}{3}} = 0{\cdot}793\,701$ ($k\theta/\epsilon_0 = 0{\cdot}406\,505$). A detailed investigation has been made of the energy and specific heat variation corresponding to the critical value of the ratio $k\theta'/\epsilon_0$ for complete saturation at absolute zero. (For $k\theta'/\epsilon_0 < 2^{-\frac{1}{3}}$, $\zeta_0 < 1$.) The results are sufficient to indicate the general

TABLE I. MAGNETIZATION, ENERGY, AND SPECIFIC HEAT AS FUNCTIONS OF TEMPERATURE IN THE LIMIT $\epsilon_0/k\theta \to 0$, CORRESPONDING TO CLASSICAL STATISTICS

θ, Curie temperature; ζ, relative magnetization; E_M, energy associated with magnetization; C_M, magnetic specific heat.

T/θ	ζ	$-E_M/Nk\theta$	C_M/Nk
0·00	1·000 000	0·500 000	0·000 000
0·05	1·000 000	0·500 000	0·000 000
0·10	1·000 000	0·500 000	0·000 001
0·15	0·999 997	0·499 997	0·000 288
0·20	0·999 909	0·499 909	0·004 547
0·25	0·999 326	0·499 326	0·021 659
0·30	0·997 414	0·497 417	0·058 100
0·35	0·993 164	0·493 187	0·114 162
0·40	0·985 624	0·485 727	0·186 636
0·45	0·973 976	0·474 315	0·271 664
0·50	0·957 504	0·458 407	0·365 948
0·55	0·935 529	0·437 607	0·466 988
0·60	0·907 332	0·411 626	0·572 977
0·65	0·872 065	0·380 249	0·682 626
0·70	0·828 635	0·343 318	0·795 016
0·75	0·775 516	0·300 713	0·909 488
0·80	0·710 412	0·252 342	1·025 560
0·85	0·629 501	0·198 136	1·142 880
0·90	0·525 430	0·138 038	1·261 185
0·95	0·379 485	0·072 005	1·380 275
1·00	0·000 000	0·000 000	1·500 000

character of the variation for the whole range of values of $k\theta'/\epsilon_0$. It may be mentioned that for greater or smaller values of $k\theta'/\epsilon_0$ the necessary preliminary interpolations to give ζ at equal intervals of T/θ could be more easily carried out and the process of interpolation would be facilitated by the detailed results available for the particular critical value.

The results given in Table II have been obtained by the method of computation outlined in § 2. In this and other computational work involved in the paper, extensive use has been made of the interpolation tables of Comrie (1936) and of the methods described in connexion with them. The starting point is Table III of I in which are given the values of T/θ at equal intervals of ζ. For a particular value of T/θ a preliminary estimate of ζ was first obtained, by approximate inverse interpolation, or by the use of series. An accurate value was found by the trial method described in connexion with (2·19). To ensure internal consistency, throughout the calculations the value taken for $k\theta/\epsilon_0$ was 0·406505, which is accurate to about 1 unit in the sixth place. For the same reason use was made of a more accurate value of

$k\theta'/\epsilon_0$ ($2^{-\frac{1}{3}} = 0{\cdot}793\ 700\ 53$), and the value of ζ was adjusted to give agreement with this value to the seventh decimal place. When ζ is known, the value of $E_M/N\epsilon_0$ is readily obtained (see 2·16). For the calculation of the distributional part of the energy, E_E, $F_{\frac{1}{2}}(\eta+\beta)$ and $F_{\frac{1}{2}}(\eta-\beta)$ are first found from (2·19), and thence the values of $\eta+\beta$ and $\eta-\beta$ by inverse interpolation in the $F_{\frac{1}{2}}(\eta)$ table. Direct interpolation in the $F_{\frac{3}{2}}(\eta)$ table then gives $F_{\frac{3}{2}}(\eta+\beta)$ and $F_{\frac{3}{2}}(\eta-\beta)$, and $E_E/N\epsilon_0$ can be calculated (see 2·16). The range of values of $\eta+\beta$ and $\eta-\beta$ required is covered by the Fermi-Dirac tables (extending from $\eta = -4$ to $\eta = +20$) except for $T/\theta = 0{\cdot}05$, $0{\cdot}10$ and $0{\cdot}15$; here $\eta+\beta > 20$, and it is necessary to make use of the asymptotic series expansions in

TABLE II. MAGNETIZATION, ENERGY, AND SPECIFIC HEAT AS FUNCTIONS OF TEMPERATURE FOR $k\theta'/\epsilon_0 = 2^{-\frac{1}{3}} = 0{\cdot}793\ 701$ ($k\theta/\epsilon_0 = 0{\cdot}406\ 505$)

$k\theta'$, interaction energy coefficient; θ, Curie temperature; ζ, relative magnetization; ϵ_0, maximum Fermi zero point energy, without interaction; E, total energy; E_M, quasi-magnetic interaction energy; E_E, energy corresponding to distribution; C, specific heat; C_M, C_E, partial specific heats corresponding to E_M, E_E. $E = E_M + E_E$. $C = C_M + C_E$. The negative sign prefixed to C_E applies only up to $T/\theta = 1{\cdot}00_-$.

T/θ	ζ	$-E_M/N\epsilon_0$	$E_E/N\epsilon_0$	$E/N\epsilon_0$	C_M/Nk	$-C_E/Nk$	C/Nk
0·00	1·00000	0·39685	0·95244	0·55559	0·0000	0·0000	0·0000
0·05	0·99665	0·39420	0·95048	0·55629	0·2057	0·1371	0·0706
0·10	0·98996	0·38892	0·94740	0·55848	0·3113	0·1650	0·1462
0·15	0·98065	0·38164	0·94391	0·56227	0·4038	0·1766	0·2272
0·20	0·96888	0·37254	0·94028	0·56774	0·4907	0·1791	0·3116
0·25	0·95471	0·36171	0·93667	0·57496	0·5740	0·1752	0·3989
0·30	0·93806	0·34921	0·93319	0·58398	0·6558	0·1671	0·4887
0·35	0·91886	0·33506	0·92990	0·59484	0·7367	0·1563	0·5804
0·40	0·89694	0·31927	0·92685	0·60758	0·8175	0·1439	0·6736
0·45	0·87210	0·30183	0·92405	0·62222	0·8984	0·1307	0·7676
0·50	0·84408	0·28275	0·92153	0·63879	0·9795	0·1175	0·8620
0·55	0·81254	0·26201	0·91927	0·65726	1·0610	0·1048	0·9562
0·60	0·77704	0·23961	0·91726	0·67765	1·1429	0·0930	1·0499
0·65	0·73698	0·21555	0·91548	0·69994	1·2251	0·0824	1·1428
0·70	0·69158	0·18981	0·91391	0·72410	1·3077	0·0730	1·2346
0·75	0·63968	0·16239	0·91251	0·75012	1·3904	0·0650	1·3254
0·80	0·57953	0·13329	0·91126	0·77797	1·4733	0·0584	1·4148
0·85	0·50821	0·10250	0·91012	0·80763	1·5563	0·0533	1·5031
0·90	0·42005	0·07002	0·90908	0·83906	1·6394	0·0493	1·5901
0·95	0·30058	0·03585	0·90811	0·87226	1·7224	0·0466	1·6758
$1{\cdot}00_-$	0·00000	0·00000	0·90718	0·90718	1·8056	0·0451	1·7605
$1{\cdot}00_+$			0·90718			1·1860	
1·05			0·93147			1·2042	
1·10			0·95612			1·2208	
1·15			0·98109			1·2359	
1·20			1·00635			1·2498	

carrying out the calculations. Accurate calculations of energy for low temperatures are therefore particularly troublesome—the difficulty would be general for other values of $k\theta'/\epsilon_0$—though the limiting value of the energy at absolute zero is readily obtained from the expressions of § 4 (ii).

The accuracy attainable in the $E/N\epsilon_0$ values varies considerably over the range covered by the table, but the uncertainty should nowhere exceed about 2 units in the sixth place, so that the tabulated five place values are subject only to rounding errors.

The reduced specific heat values have been obtained by numerical differentiation of the energy values, giving $d(E/N\epsilon_0)/d(T/\theta)$, followed by multiplication by $\epsilon_0/k\theta$ ($\epsilon_0/k\theta = 2{\cdot}45999$), giving C/Nk (see 2·18). It is difficult to make precise estimates of the accuracy of derivatives found by numerical differentiation of a tabulated function. Six-place values of $E/N\epsilon_0$ were used, and an examination of the sequence of contributions from successive differences indicates that the rounded values given in Table II for C/Nk are probably correct to about 1 unit in the fourth place throughout. This degree of accuracy also holds for the C_M/Nk and C_E/Nk values for $T/\theta \geqslant 0{\cdot}2$, but in the initial part of the range, where the slowly converging forward difference formula must be used, and the higher order differences remain rather large, the uncertainties are greater; the greatest uncertainty is in the values for $T/\theta = 0{\cdot}05$, where it may amount to about 4 units in the fourth place.

For $T/\theta > 1$, for each value of T/θ the corresponding value of kT/ϵ_0 is first obtained by multiplication by $k\theta/\epsilon_0$, and $E/N\epsilon_0$ and C/Nk found from interpolation in the electronic energy and specific heat tables (Stoner 1938*c*).

The numerical results of Table II are shown graphically in figs. 1 and 2.

The peculiarities of the results, as compared with those given by classical statistics, are brought out very clearly by these curves. The variation of the quasi-magnetic part of the interaction energy, E_M, has the same general character, except at low temperatures where the rate of variation is greater, as shown more clearly in the corresponding curve for the partial specific heat, C_M. The most striking difference from the classical relations lies in the fact that the distributional part of the electronic energy, E_E, varies discontinuously at the Curie point as well as E_M. For this particular value of $k\theta'/\epsilon_0$, the discontinuity in C_E at the Curie point ($\Delta C_E/Nk = +1{\cdot}2311$) is comparable with that of C_M ($\Delta C_M/Nk = -1{\cdot}8056$), and in the opposite sense, so that the discontinuity in the total electronic specific heat, C, ($\Delta C/Nk = -0{\cdot}5745$) is very misleading as an indication of the magnitude of the quasi-magnetic interaction effect. Again, for this value of $k\theta'/\epsilon_0$, the

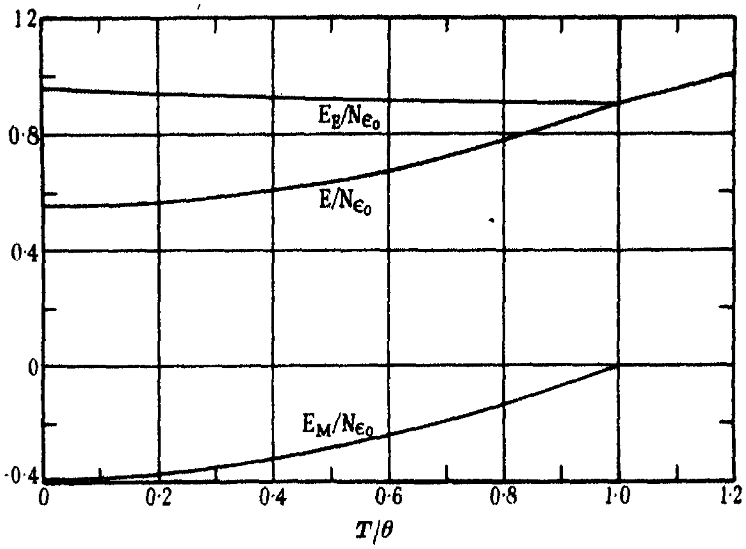


FIG. 1. Electronic energy as a function of temperature for $k\theta'/\epsilon_0 = 2^{-\frac{1}{2}}$. For symbols, see Table II.

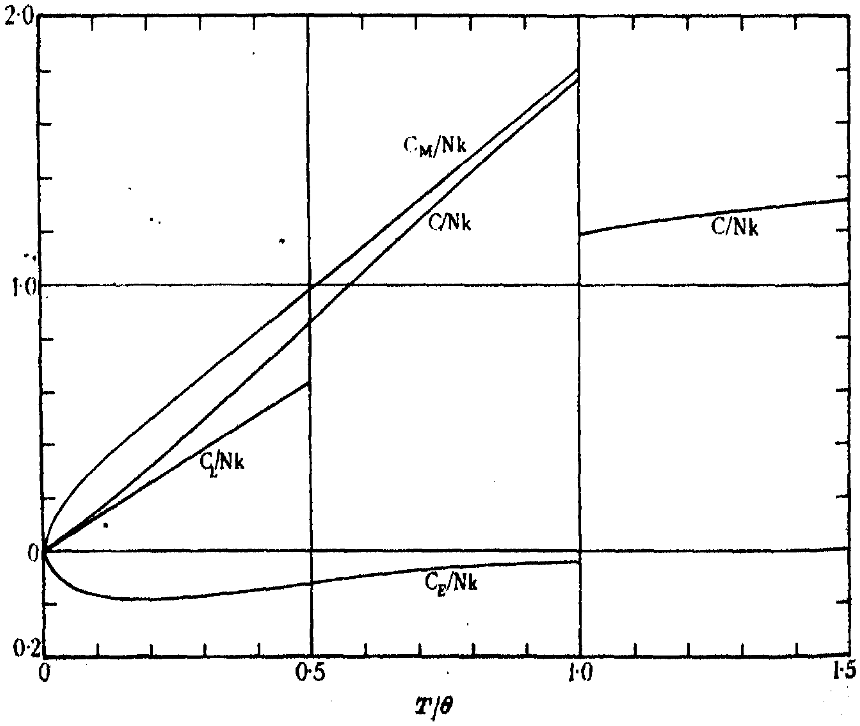


FIG. 2. Electronic specific heat as a function of temperature for $k\theta'/\epsilon_0 = 2^{-\frac{1}{2}}$. For symbols, see Table II. The straight line C_L/Nk corresponds to the limiting low temperature expression (4·22) for the specific heat.

increase in the quasi-magnetic energy between absolute zero and the Curie point ($\Delta E_M/N\epsilon_0 = 0{\cdot}39685$) is actually greater than the increase in total electronic energy ($\Delta E/N\epsilon_0 = 0{\cdot}35159$).

The labour entailed in detailed calculations of the energy and specific heat variation for a wide range of values of $k\theta'/\epsilon_0$ would hardly be justified in view of the limited application which can be made of them. It is sufficient to calculate some of the salient magnitudes characterizing the energy and specific heat relations whose variation with $k\theta'/\epsilon_0$ is of interest, and which serve, in conjunction with the detailed results for $k\theta'/\epsilon_0 = 2^{-\frac{1}{2}}$, to give a reasonably complete indication of the character of the variation with temperature of the energy and specific heat for each value of $k\theta'/\epsilon_0$. The results of such calculations are given in the next section.

6. DEPENDENCE ON $k\theta'/\epsilon_0$ OF ENERGY AND SPECIFIC HEAT CHARACTERISTICS

An examination of figs. 1 and 2 indicates that the energy and specific heat relations for any value of $k\theta'/\epsilon_0$ may be briefly characterized by the values of the energy and specific heat at low temperatures and at the Curie point. The quasi-magnetic and distributional parts of the energy, E_M and E_E, and their sum, the total energy, E, at absolute zero are readily obtained from (4·12). At the Curie point, the quasi-magnetic energy is zero, and $E = E_E$. The Curie point energy, E_θ, is most conveniently obtained by interpolation in the electronic energy tables, giving $E/N\epsilon_0$ as a function of kT/ϵ_0 (Stoner 1938*c*). From an experimental point of view, energy differences between absolute zero and the Curie point are more significant than the energy values at these two temperatures. Instead of tabulating the ratios $\Delta E/N\epsilon_0$, however, it is appropriate to tabulate $\Delta E/Nk\theta$, as being of more direct experimental interest. Moreover, the relation with classical results is brought out more clearly by these ratios. The results of the calculations are shown in Table III. Following the procedure of I, it is convenient to take ζ_0, the relative magnetization at absolute zero, rather than $k\theta'/\epsilon_0$, as the argument in the tables for $k\theta'/\epsilon_0 \leqslant 2^{-\frac{1}{2}}$. As $k\theta'/\epsilon_0$ increases the ratios $(E_\theta - E_0)_M/Nk\theta$, $(E_\theta - E_0)_E/Nk\theta$ and $(E_\theta - E_0)/Nk\theta$ approach asymptotically the classical limiting values of 0·5, 1·5 and 2·0 respectively.

The ratios are shown in Fig. 3 as functions of the reduced Curie temperature $k\theta/\epsilon_0$. The curves bring out clearly the remarkable differences in the relative contributions from the magnetic and distributional energy to the total energy increase for different values of $k\theta/\epsilon_0$, and the wide divergence of the ratios from the classical values. It may be noted that for small values

TABLE III. ENERGY CHARACTERISTICS FOR RANGE OF VALUES OF $k\theta'/\epsilon_0$

The table gives energies at absolute zero and at the Curie point, and energy differences, as functions of ζ_0 for $k\theta'/\epsilon_0 \leqslant 2^{-\frac{1}{3}}$, and of $k\theta'/\epsilon_0$ for $k\theta'/\epsilon_0 > 2^{-\frac{1}{3}}$. ζ_0, relative magnetization at absolute zero. For other symbols, see Table II. For $k\theta'/\epsilon_0 \geqslant 2^{-\frac{1}{3}} = 0{\cdot}79370$, $\zeta_0 = 1$, $(E_E)_0/N\epsilon_0 = 0{\cdot}95244$.

ζ_0	$\frac{k\theta'}{\epsilon_0}$	$\frac{k\theta}{\epsilon_0}$	$-\frac{(E_M)_0}{N\epsilon_0}$	$\frac{(E_E)_0}{N\epsilon_0}$	$\frac{E_0}{N\epsilon_0}$	$\frac{E_\theta}{N\epsilon_0}$	$\frac{(E_\theta-E_0)_M}{Nk\theta}$	$\frac{(E_\theta-E_0)_E}{Nk\theta}$	$\frac{E_\theta-E_0}{Nk\theta}$
0·0	0·66667	0·00000	0·00000	0·60000	0·60000	0·60000	0·00000	0·00000	0·00000
0·1	0·66716	0·03001	0·00334	0·60333	0·60000	0·60222	0·11116	−0·03718	0·07398
0·2	0·66867	0·05999	0·01337	0·61335	0·59998	0·60883	0·22292	−0·07538	0·14755
0·3	0·67127	0·08985	0·03021	0·63010	0·59989	0·61967	0·33619	−0·11612	0·22007
0·4	0·67511	0·11938	0·05401	0·65366	0·59965	0·63435	0·45239	−0·16179	0·29060
0·5	0·68041	0·14878	0·08505	0·68416	0·59911	0·65253	0·57167	−0·21260	0·35907
0·6	0·68758	0·17891	0·12376	0·72178	0·59801	0·67445	0·69176	−0·26453	0·42723
0·7	0·69733	0·21130	0·17085	0·76678	0·59593	0·70119	0·80853	−0·31042	0·49811
0·8	0·71108	0·24856	0·22755	0·81957	0·59203	0·73533	0·91544	−0·33891	0·57654
0·9	0·73255	0·29682	0·29668	0·88086	0·58418	0·78389	0·99954	−0·32670	0·67284
0·95	0·75006	0·33120	0·33846	0·91512	0·57666	0·82088	1·02194	−0·28456	0·73738
1·00	0·79370	0·40650	0·39685	0·95244	0·55559	0·90718	0·97625	−0·11134	0·86491
	0·8	0·41657	0·40	0·95244	0·55244	0·91916	0·96023	−0·07989	0·88034
	0·9	0·56151	0·45	—	0·50244	1·10008	0·80141	+0·26294	1·06435
	1·0	0·69116	0·50	—	0·45244	1·27106	0·72342	0·46099	1·18441
	1·1	0·81316	0·55	—	0·40244	1·43685	0·67638	0·59571	1·27209
	1·2	0·93052	0·60	—	0·35244	1·59942	0·64480	0·69529	1·34009
	1·3	1·04477	0·65	—	0·30244	1·75980	0·62214	0·77276	1·39491
	1·4	1·15679	0·70	—	0·25244	1·91860	0·60512	0·83520	1·44033
	1·5	1·26714	0·75	—	0·20244	2·07620	0·59188	0·88684	1·47873
	1·6	1·37618	0·80	—	0·15244	2·23285	0·58132	0·93041	1·51173
	1·7	1·48419	0·85	—	0·10244	2·38876	0·57270	0·96775	1·54045
	1·8	1·59134	0·90	—	0·05244	2·54405	0·56556	1·00017	1·56573
	1·9	1·69780	0·95	—	0·00244	2·69882	0·55955	1·02862	1·58816
	2·0	1·80366	1·00	—	−0·04756	2·85317	0·55443	1·05382	1·60824

of $k\theta/\epsilon_0$ the increase in the quasi-magnetic part of the interaction energy exceeds the total energy increase. The critical value of $k\theta/\epsilon_0$ for which the increase in distributional energy is zero corresponds to the value of kT/ϵ_0 for which the normal $E/N\epsilon_0$ is equal to that for complete saturation at absolute zero as given by (4·13). Inverse interpolation in the electronic energy table gives the critical value as 0·444 138.

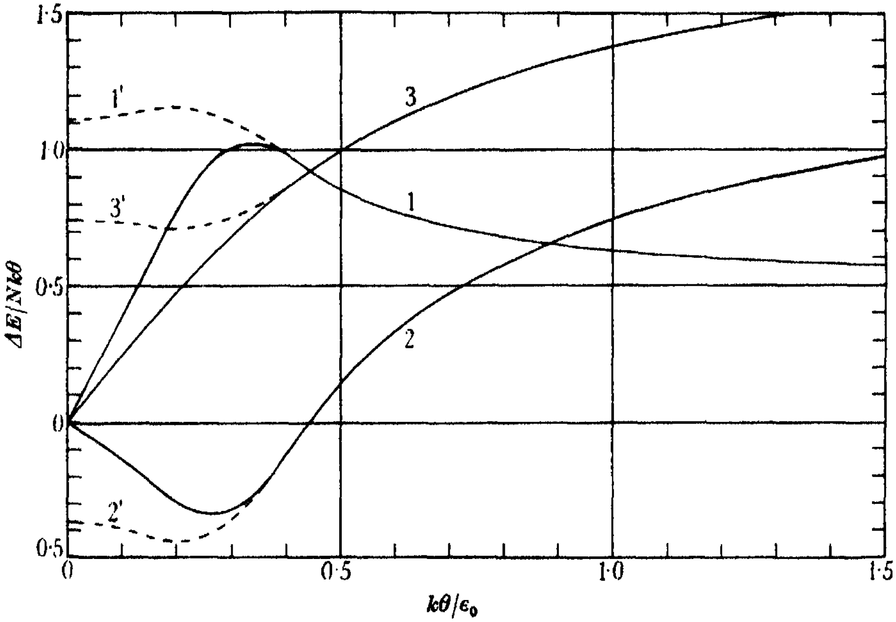


FIG. 3. Energy increase between absolute zero and the Curie temperature as a function of $k\theta/\epsilon_0$. 1, $(E_\theta - E_0)_M/Nk\theta$; 2, $(E_\theta - E_0)_E/Nk\theta$; 3, $(E_\theta - E_0)/Nk\theta$. The broken curves 1′, 2′, 3′ give the ratios 1, 2, 3 divided by ζ_0. E_M, quasi-magnetic interaction energy; E_E, energy corresponding to distribution; E, total energy; θ, Curie temperature; ϵ_0, maximum Fermi zero point energy, without interaction; ζ_0, relative magnetization at absolute zero.

The specific heat characteristics at low temperatures and at the Curie point are given in Table IV. The limiting value of the ratio $(C/Nk)/(kT/\epsilon_0)$ for $kT/\epsilon_0 \to 0$ is obtained from (4·22). Multiplication by $k\theta/\epsilon_0$ gives the limiting value of the ratio $(C/Nk)/(T/\theta)$ for $T/\theta \to 0$. The closeness of approximation given by this low temperature expression for the specific heat in the low temperature range is illustrated by the curves of fig. 2 for $k\theta'/\epsilon_0 = 2^{-\frac{1}{2}}$. It does not seem possible to obtain convenient general expressions for the partial specific heats, C_M and C_E at low temperatures. The total and partial specific heats at the Curie point (at $\theta - 0$) are calculated from the expressions

(4·6). The specific heat immediately above the Curie point (at $\theta + 0$) is given by the expression for C_1 in (4·6), and was calculated from this expression, and also by the method of direct interpolation in the electronic specific heat tables. The values of $(C/Nk)_{\theta+0}$ are reliable to 1 unit in the fourth place throughout. The evaluation of the specific heat at $\theta - 0$ by means of (4·6) involves the calculation of $F_{\frac{1}{2}}(\eta)$ from (2·2), the determination of η by inverse

TABLE IV. SPECIFIC HEAT CHARACTERISTICS FOR RANGE OF VALUES OF $k\theta'/\epsilon_0$

The table gives the limiting low temperature ratio of the specific heat to the temperature, the partial and total specific heats just below, and the total specific heat just above the Curie point, and the specific heat discontinuity. For values of $k\theta'/\epsilon_0$ and $k\theta/\epsilon_0$ corresponding to the tabulated ζ_0, and of $k\theta/\epsilon_0$ corresponding to the tabulated $k\theta'/\epsilon_0$, see Table III. See Tables II and III for symbols. For $\zeta_0 = 1$ ($k\theta'/\epsilon_0 \geqslant 0\cdot79370$), the limiting low temperature ratio $(C/Nk)/(kT/\epsilon_0)$ has the constant value $2^{\frac{1}{3}}\pi^2/4 = 3\cdot1087$.

	$kT/\epsilon_0 \to 0$	$T/\theta \to 0$	$T = \theta - 0$			$T = \theta + 0$	
ζ_0	$\dfrac{C/Nk}{kT/\epsilon_0}$	$\dfrac{C/Nk}{T/\theta}$	$\dfrac{C_M}{Nk}$	$\dfrac{C_E}{Nk}$	$\dfrac{C}{Nk}$	$\dfrac{C}{Nk}$	$\dfrac{\Delta C}{Nk}$
0·0	4·9348	0·0000	0·000	0·000	0·000	0·0000	0·000
0·1	4·9293	0·1479	0·222	−0·074	0·148	0·1477	0·000
0·2	4·9125	0·2947	0·446	−0·151	0·295	0·2928	0·003
0·3	4·8837	0·4388	0·67	−0·23	0·44	0·4319	0·01
0·4	4·8413	0·5780	0·90	−0·32	0·58	0·5601	0·02
0·5	4·7828	0·7116	1·15	−0·43	0·72	0·6749	0·05
0·6	4·7039	0·8416	1·35	−0·48	0·87	0·7773	0·09
0·7	4·5966	0·9713	1·52	−0·49	1·02	0·8707	0·15
0·8	4·4444	1·1047	1·651	−0·459	1·192	0·9591	0·233
0·9	4·2013	1·2470	1·747	−0·355	1·392	1·0494	0·343
0·95	3·9916	1·3220	1·781	−0·261	1·520	1·1007	0·419
1·00	3·1087	1·2637	1·806	−0·045	1·761	1·1860	0·575
$k\theta'/\epsilon_0$							
0·8	3·1087	1·2950	1·805	−0·017	1·788	1·1952	0·598
0·9	—	1·7456	1·773	+0·336	2·109	1·2915	0·818
1·0	—	2·1486	1·733	+0·566	2·300	1·3422	0·957
1·1	—	2·5279	1·699	0·727	2·426	1·3738	1·053
1·2	—	2·8927	1·673	0·845	2·517	1·3955	1·122
1·3	—	3·2479	1·651	0·934	2·586	1·4113	1·174
1·4	—	3·5962	1·633	1·005	2·638	1·4233	1·215
1·5	—	3·9392	1·619	1·061	2·680	1·4327	1·248
1·6	—	4·2782	1·608	1·107	2·715	1·4403	1·275
1·7	—	4·6139	1·597	1·146	2·743	1·4465	1·296
1·8	—	4·9471	1·589	1·178	2·767	1·4517	1·315
1·9	—	5·2780	1·582	1·206	2·788	1·4560	1·331
2·0	—	5·6071	1·575	1·230	2·805	1·4598	1·345

interpolation in the $F_{\frac{1}{2}}(\eta)$ table, and the subsequent determination of $F_{\frac{1}{2}}(\eta)$, $F'(\eta)$, $F''(\eta)$ and $F'''(\eta)$ (and also $F(\eta)$ as a check) by direct interpolation in the corresponding tables. The process is not unduly lengthy, particularly in view of the labour which would be entailed if the alternative method of calculating energies for a number of temperatures near the Curie temperature were to be applied. The accuracy, however, is limited by that of $F'''(\eta)$, the least accurately known of the functions entering into the formulae. It is unnecessary to describe in detail how the uncertainty in the value of $F'''(\eta)$ affects the final results. For values of $k\theta/\epsilon_0$ greater than that corresponding to $\zeta_0 = 0{\cdot}7$, the values of C_M/Nk, C_E/Nk and C/Nk at $\theta-0$ are probably correct to about 1 unit in the third decimal place. For smaller values of $k\theta/\epsilon_0$, however, the η range corresponds to $\eta > 4$, for which the Fermi Dirac functions have been evaluated to a lower absolute accuracy (to one decimal place less), and there is an uncertainty of about 1 unit in the second decimal place in the specific heat values. This uncertainty would increase for the lowest range of $k\theta/\epsilon_0$ values, but here, for $\zeta_0 \leqslant 0{\cdot}2$, a higher accuracy can be obtained by the use of the low temperature formulae (4·9). The limiting values of the ratios $(C_M/Nk)_{\theta-0}$, $(C_E/Nk)_{\theta-0}$, $(C/Nk)_{\theta-0}$, $(C/Nk)_{\theta+0}$ and $(\Delta C/Nk)$ for $\epsilon_0/k\theta' \to 0$ are 1·5, 1·5, 3·0, 1·5 and 1·5 respectively.

The reduced values of the partial and total specific heats, and of the discontinuity at the Curie point are shown in fig. 4 as functions of $k\theta/\epsilon_0$. The curve 4, $(C/Nk)_{\theta+0}$, is the usual electronic specific heat curve. The curves bring out the remarkable variation in the specific heat relations for small values of $k\theta/\epsilon_0$, and also the approach to the classical relations as $k\theta/\epsilon_0$ increases.

The limiting low temperature specific heat ratios are shown graphically in fig. 5 as functions of $k\theta/\epsilon_0$. The corresponding values of ζ_0 are indicated on the top scale. The discontinuous variation at $k\theta/\epsilon_0 = 0{\cdot}4065$ arises from the fact that at higher values of $k\theta/\epsilon_0$ the electron spins are all parallel at absolute zero; as $k\theta/\epsilon_0$ decreases through the critical value there is a contribution to the specific heat from the anti-parallel group of spins which at first increases very rapidly, as $(1-\zeta_0)^{\frac{1}{2}}$ (see 4·15).

The quantities which are, in principle, experimentally determinable for a ferromagnetic, are the electronic specific heat (including the values at low temperatures and at the Curie point) and electronic energy differences (including the energy difference between absolute zero and the Curie point). The tabulated ratios, such as $(E_\theta - E_0)/Nk\theta$ and $(C/Nk)_{\theta-0}$, involve the number, N, of potentially effective electrons, which is not known directly from experiment. If the energy and specific heat are referred to the gram atom, $N = N_A q$, where N_A is Avogadro's number, q the number of effective

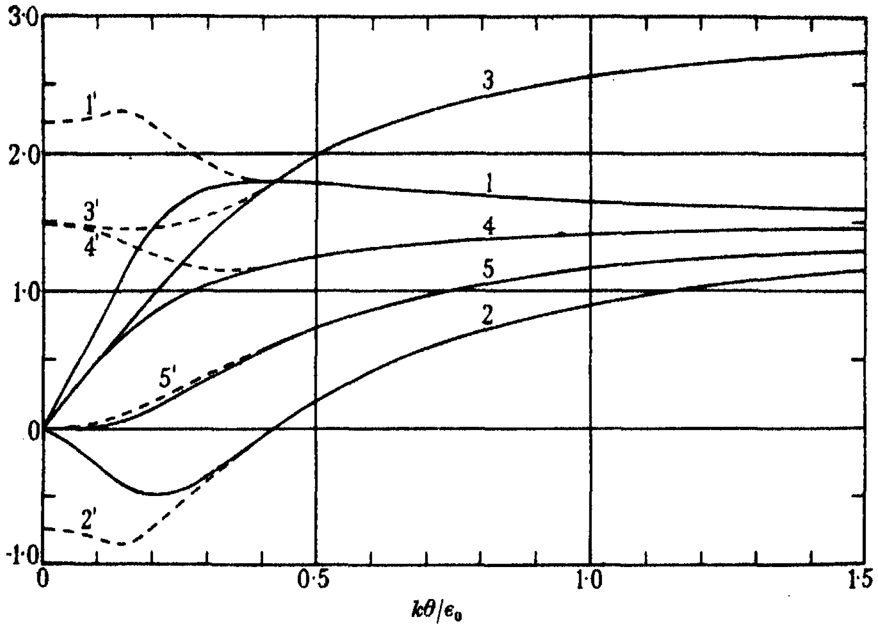


FIG. 4. Total and partial specific heats, and specific heat discontinuity at the Curie point, as functions of $k\theta/\epsilon_0$. 1, $(C_M/Nk)_{\theta-0}$; 2, $(C_E/Nk)_{\theta-0}$; 3, $(C/Nk)_{\theta-0}$; 4, $(C/Nk)_{\theta+0}$; 5, $\Delta C/Nk$. The broken curves 1′, 2′, 3′, 4′, 5′ give the ratios 1, 2, 3, 4, 5 divided by ζ_0. C_M, quasi-magnetic specific heat; C_E, distributional specific heat; C, total specific heat ($C = C_E + C_M$); θ, Curie temperature.

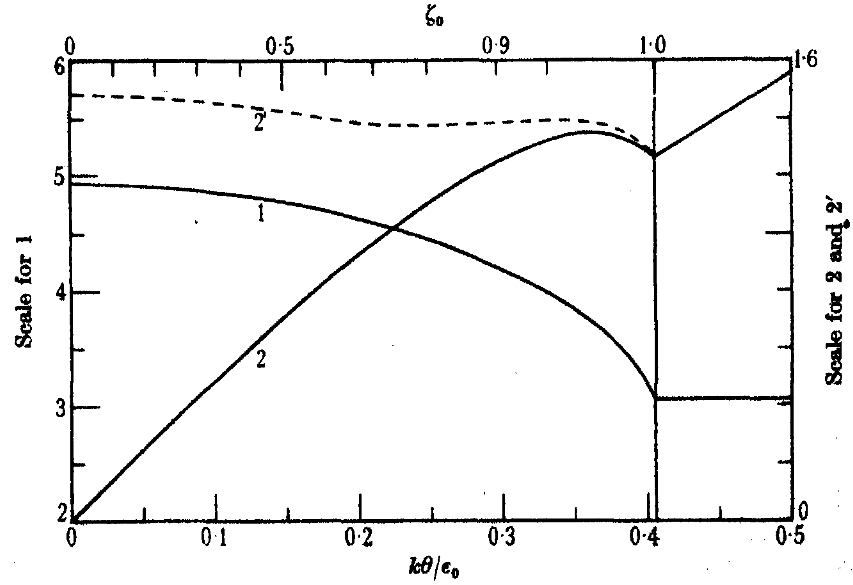


FIG. 5. Limiting low temperature ratio of specific heat to temperature as a function of $k\theta/\epsilon_0$.

1, $\{(C/Nk)/(kT/\epsilon_0)\}_{kT/\epsilon_0\to 0}$; 2, $\{(C/Nk)/(T/\theta)\}_{T/\theta\to 0}$; 2′, $\{(C/N\zeta_{00}k)/(T/\theta)\}_{T/\theta\to 0}$.

electrons per atom. The value which is determinable from experiment is that of $q\zeta_0$, which is given directly by the saturation magnetization at low temperatures. For nickel, for example, $q\zeta_0 = 0{\cdot}606$. To reduce the various tabulated ratios to a form involving experimentally determinable quantities, division by ζ_0 is necessary; this changes the ratios only in the range $k\theta'/\epsilon_0 < 2^{-\frac{1}{3}}$ for which $\zeta_0 < 1$. The modified ratios are readily obtained from Tables III and IV. It is unnecessary to give all of them, but they are shown by the broken curves of figs. 3–5. The values of the ratios involving directly determinable quantities are tabulated in Table V. For higher values of $k\theta'/\epsilon_0$, where $\zeta_0 = 1$, the ratios continue as in the corresponding columns of Tables III and IV. The limiting values of the ratios for $\zeta_0 \to 0$ (given in brackets in Table V) are found by making use of the relation (3·9),

$$\frac{k\theta}{\epsilon_0} \to \frac{2\sqrt{2}}{3\pi}\zeta_0 \quad \text{for} \quad \zeta_0 \to 0,$$

in combination with the low temperature limiting formulae of § 4. They are set out below.

For $\zeta_0 \to 0$ $(k\theta'/\epsilon_0 \to 0)$:

$$\frac{(E_\theta - E_0)_M}{N\zeta_0 k\theta} \to \frac{\pi\sqrt{2}}{4}; \quad \frac{(E_\theta - E_0)_E}{N\zeta_0 k\theta} \to -\frac{\pi\sqrt{2}}{12}; \quad \frac{E_\theta - E_0}{N\zeta_0 k\theta} \to \frac{\pi\sqrt{2}}{6}; \tag{6·1}$$

$$\frac{(C_M)_{\theta-0}}{N\zeta_0 k} \to \frac{\pi\sqrt{2}}{2}; \quad \frac{(C_E)_{\theta-0}}{N\zeta_0 k} \to -\frac{\pi\sqrt{2}}{6}; \quad \frac{C_{\theta-0}}{N\zeta_0 k} \to \frac{\pi\sqrt{2}}{3}; \quad \frac{C_{\theta+0}}{N\zeta_0 k} \to \frac{\pi\sqrt{2}}{3}. \tag{6·2}$$

To this approximation, $\Delta C/N\zeta_0 k$ tends to zero. To the next approximation,

$$\frac{\Delta C}{N\zeta_0 k} \to \frac{\pi^3\sqrt{2}}{12}\left(\frac{k\theta}{\epsilon_0}\right)^2 = \frac{2\pi\sqrt{2}}{27}\zeta_0^2. \tag{6·3}$$

For computational purposes it may be noted that

$$\pi\sqrt{2} = 4{\cdot}442\,883.$$

It may be noticed that none of the experimentally determinable ratios of Table V varies considerably over the $k\theta/\epsilon_0$ range for which $\zeta_0 < 1$; these ratios will not therefore provide a sensitive means of estimating ζ_0. The energy difference ratio, however, subsequently increases fairly rapidly (see fig. 3, curves 3, 3′) so that from this a fairly definite indication can be obtained of whether the value of $k\theta/\epsilon_0$ exceeds the critical value for complete saturation; a similar indication can be obtained from the low temperature specific heat ratio (see Tables IV and V). In principle the most definite criterion of the

TABLE V. EXPERIMENTALLY DETERMINABLE ENERGY AND SPECIFIC HEAT CHARACTERISTICS FOR THE RANGE $k\theta'/\epsilon_0 \leqslant 2^{-\frac{1}{2}}$

See Table III for values of $k\theta'/\epsilon_0$ and $k\theta/\epsilon_0$ corresponding to the tabulated ζ_0, and for symbols.

ζ_0	$\dfrac{E_\theta - E_0}{N\zeta_0 k\theta}$	$\left(\dfrac{C/N\zeta_0 k}{T/\theta}\right)_{T/\theta\to 0}$	$\dfrac{C_{\theta-0}}{N\zeta_0 k}$	$\dfrac{C_{\theta+0}}{N\zeta_0 k}$
0·0	(0·74048)	(1·4810)	(1·481)	(1·481)
0·1	0·73980	1·4792	1·480	1·477
0·2	0·73773	1·4735	1·478	1·464
0·3	0·73358	1·4627	1·47	1·440
0·4	0·72651	1·4450	1·46	1·400
0·5	0·71814	1·4232	1·46	1·350
0·6	0·71205	1·4026	1·45	1·296
0·7	0·71159	1·3875	1·46	1·244
0·8	0·72067	1·3809	1·490	1·199
0·9	0·74759	1·3856	1·547	1·166
0·95	0·77619	1·3916	1·600	1·158
1·00	0·86491	1·2637	1·761	1·186

$k\theta/\epsilon_0$ value, at least in the lower range, appears to be the magnitude of the specific heat discontinuity (see fig. 4, curves 5, 5′), though here there may be difficulties on the experimental side. It should be remembered in connexion with all these computational results, that the detailed character of the relations found is necessarily dependent on the particular form assumed for the energy distribution of states; the determination of the relations which hold for a particular band form may, however, be of great value in the analysis of experimental results, even if the band form in actual ferromagnetics differs widely from that assumed. It seems desirable to develop the whole series of interconnected relations if only to prevent erroneous conclusions being drawn from the application of one type of relation to experimental results.

7. GENERAL DISCUSSION

The general remarks made in the discussion in I on the character of the collective electron treatment of the magnetic characteristics of ferromagnetics are applicable, with but slight modification, to the treatment of the energetic characteristics. Sets of relations have been obtained which are consistent with the premises involved in the initial equations, and an analysis of the experimental results should provide evidence as to how far those premises are appropriate as an approximate representation of conditions in ferromagnetics. Owing to the fact that the theory is concerned with the spontaneous magnetization below the Curie point, whose determination, from magnetic data alone, may be subject to considerable uncertainty, an examination of the associated energetic characteristics is essential for a survey of the implications of the treatment which is adequate to form a proper basis for the consideration of experimental results.

With a classical treatment, at least in its simpler form, both the magnetization and energy relations over the whole temperature range are completely determined if the saturation magnetization at absolute zero and the Curie temperature are known. With a Fermi-Dirac treatment, a complete specification requires in addition a knowledge of the form of the energy distribution of electronic energy states, and of the Fermi zero point energy. For a particular type of distribution, the additional knowledge required reduces to that of the Fermi zero point energy. With the present treatment, therefore, for complete specification the values are required of the saturation magnetization (which gives the product, $q\zeta_0$, of the number of potentially effective electrons per atom and the relative magnetization at absolute zero), the Curie temperature, θ, and the Fermi energy, ϵ_0. Taking $q\zeta_0$ and θ as known, an estimate of ϵ_0 can, in principle, be made from a knowledge of the electronic specific heat at any temperature, or from a knowledge of electronic energy differences, as well as from the spontaneous magnetization at temperatures below, and from the paramagnetism at temperatures above the Curie point. Even when the experimental specific heat results are sufficiently accurate, there are, in general, considerable uncertainties in the estimates of the electronic part of the specific heat, owing to various difficulties in the analysis of the results. Assuming that reasonably accurate estimates of the electronic specific heat have been made, more particularly at low temperatures and in the Curie point region, and of energy differences between absolute zero and the Curie point, agreement between the values of ϵ_0 deduced in different ways would indicate that the basic assumptions made as to band form in the treatment were satisfactory as a representation of the physical conditions in the actual ferromagnetic. Even considerable differences in the effective values of ϵ_0 deduced would not, however, necessarily be inconsistent with the general point of view underlying the treatment, but would rather provide a starting point for a more complete theoretical treatment of the properties of the particular material. As a discussion of the experimental results will be taken up later in the light of the treatments of both the magnetic and the energetic characteristics, it will be sufficient to give here a very brief preliminary survey of the experimental results for nickel in order to indicate the kind of relation which holds between the theoretical results of the present paper and the properties of actual ferromagnetics.

It has been shown (Stoner 1938*b*) that the temperature variation of the paramagnetism of nickel above the Curie point corresponds to a $k\theta'/\epsilon_0$ value very slightly less than that for complete saturation at absolute zero. The corresponding theoretical energy and specific heat relations would then approximate closely to those for $k\theta'/\epsilon_0 = 2^{-\frac{1}{3}}$, which have been considered in detail (Table II, figs. 1, 2). One of the most obvious differences in the theoretical and experimental specific heat variation lies in the fact that the experimental specific heat does not fall abruptly to a minimum at the Curie point. There is at present no entirely satisfactory explanation of the "tail" in the specific heat curve, though it is undoubtedly due to secondary effects, possibly of the local ordering type, which are entirely neglected in the present treatment. Taking the maximum specific heat as $C_{\theta-0}$ and the minimum specific heat above the Curie point at $C_{\theta+0}$, the experimental values of $C_{\theta-0}$, $C_{\theta+0}$ and ΔC_θ divided by $Nk\zeta_0$ are approximately 2·5, 1·0 and 1·5. The values of the first and third of these ratios are considerably higher than those corresponding to the critical $k\theta'/\epsilon_0$ for $\zeta_0 = 1$ (see Table V). In one sense the higher maximum specific heat is satisfactory, as being in keeping with the fact that experimentally the spontaneous magnetization falls off more rapidly as the Curie point is approached from below than the theoretical treatment would indicate, a discrepancy previously noticed (I, p. 411), and possibly arising from the neglect of domain effects.

Although the magnitude of the electronic specific heat indicated by the experimental data above and near the Curie point is somewhat lower than that indicated by the theoretical treatment, at temperatures in the neighbourhood of 1000° K the only available measurements suggest that it is somewhat higher (cf. Stoner 1938*c*). In view of the uncertainties in the analysis of the experimental results, as well as the uncertainties in the experimental results themselves, the agreement should perhaps be regarded as reasonably close. The discrepancies are of a similar order of magnitude to those found for non-ferromagnetic metals.

It has previously been shown (Stoner 1938*b*) that the high temperature paramagnetism is consistent with a value of ϵ_0 of the same order as that indicated by the low temperature electronic specific heat. There is, however, a definite discrepancy, which is brought out particularly clearly if the ratio $(C/N\zeta_0 k)/(T/\theta)$ for $T/\theta \rightarrow 0$ is considered (Table V). For nickel ($q\zeta_0 = 0{\cdot}606$, $\theta = 631$) the minimum value of this ratio ($1{\cdot}2637$ for $\zeta_0 = 1$) corresponds to a value of $2{\cdot}41 \times 10^{-3}$ for the limiting ratio of the electronic specific heat per gram atom to the temperature. The observed value is $1{\cdot}74 \times 10^{-3}$. Alternatively expressed, the observed low temperature specific heat is consistent with a Curie temperature of not less than 872° K, as compared with the observed temperature of 631° K. The discrepancy here is almost certainly due to a difference in the band form in the actual metal from that assumed in the treatment, the relations involved being much more critically dependent on the precise band form than are those for the high temperature paramagnetism and specific heat, and for electronic energy differences.

From a previous analysis of the experimental results for nickel (Stoner 1936*b*) the electronic energy difference ratio, $(E_\theta - E_0)/N\zeta_0 k\theta$, may be estimated as approximately 0·8. An inspection of Tables V and III shows that this is compatible with a $k\theta'/\epsilon_0$ ratio only slightly less than that required to give complete saturation at absolute zero. This energy difference ratio is much less dependent on secondary effects on the experimental side and on special assumptions in the theoretical treatment than are those ratios involving the value of the specific heat at a particular temperature. The experimental ratio characterizes the change in the complete range of temperature up to the Curie point, and it is satisfactory that the effective $k\theta'/\epsilon_0$ value which the ratio indicates for nickel is in close agreement with that indicated by the paramagnetism over a wide range of temperature above the Curie point.

The standard band form assumed in the treatment is probably more closely approached in nickel than in other ferromagnetics. The brief survey of the experimental results suggests that even for nickel the approximation is not sufficiently close for the results to be covered with precision over the whole temperature range by an unchanging effective value of ϵ_0. The general character of the fundamental experimental results, however, is in remarkable accordance with that suggested by the theoretical treatment. Essentially, the theoretical treatment as a whole provides interconnected sets of magnetic and energetic relations for idealized ferromagnetics of a very simple type. The discussion in this section shows that these relations are in sufficiently close quantitative agreement with those found for actual ferromagnetics to make the treatment of great potential value for the detailed analysis and interpretation of the observed characteristics of particular ferromagnetic materials.

I have pleasure in acknowledging my indebtedness to the Government Grants Committee of the Royal Society for the loan of a Brunsviga calculating machine.

SUMMARY

The treatment of collective electron ferromagnetism on the basis of Fermi-Dirac statistics, applied in a previous paper to magnetic characteristics, is

extended to energetic characteristics. The theory is developed for electrons in an energy band of standard form and subject to interaction effects giving rise to a dependence of energy on the square of the magnetization. The distribution is specified by a parameter ϵ_0, the Fermi zero point energy in the absence of this interaction. The relative magnetization, ζ, varies with temperature in a manner depending on $k\theta'/\epsilon_0$, where $k\theta'$ is a measure of the maximum quasi-magnetic interaction per electron. The Curie temperature, θ, is less than θ' except in the classical limit ($\epsilon_0/k\theta' \rightarrow 0$). In the previous paper detailed numerical results were given for the variation of ζ with T/θ for a wide range of values of $k\theta'/\epsilon_0$.

General expressions are here developed for the electronic energy, which may be analysed into a quasi-magnetic and a distributional part (§ 2). The forms taken by these expressions in the classical and Fermi Dirac limits are obtained (§ 3). General expressions are also derived for the specific heat at low temperatures and at the Curie point (§ 4).

Detailed computations are made (§ 5) of the energy and specific heat over the whole temperature range for $\epsilon_0/k\theta' \rightarrow 0$ (Table I), and for $k\theta'/\epsilon_0 = 2^{-\frac{1}{3}}$, the critical value for complete saturation ($\zeta_0 = 1$) at absolute zero (Table II, figs. 1, 2). Computations are also made (§ 6) for a range of values of $k\theta'/\epsilon_0$ (corresponding to the range from $k\theta/\epsilon_0 \rightarrow 0$ to $\epsilon_0/k\theta \rightarrow 0$) of the energy increase between absolute zero and the Curie point, and of the specific heat at low temperatures and immediately below and immediately above the Curie point (Tables III–V, figs. 3–5).

It is shown (§ 4) that the limiting low temperature ratio of the specific heat to the temperature depends only on the distribution at absolute zero, being independent of the character of the variation of spontaneous magnetization. Except in the classical limit there is a discontinuous variation in the distributional as well as in the quasi-magnetic energy at the Curie point, and an associated discontinuity in the specific heat, so that the discontinuity in the total specific heat is not an immediate measure of the quasi-magnetic interaction. Moreover, the ratio of the increase in quasi-magnetic energy between absolute zero and the Curie point to the total energy increase may diverge widely from the classical ratio. For small values of $k\theta/\epsilon_0$ it is greater than unity (Table III, fig. 3).

A preliminary survey of the specific heat results for nickel is made (§ 7). Although the observed energetic characteristics over the whole temperature range cannot be covered by an unchanging effective value of ϵ_0, the general character of the results is in good accordance with the theoretical treatment, which provides a basis for a more complete analysis and interpretation of experimental results for particular ferromagnetics.

REFERENCES

Comrie, L. J. 1936 "Interpolation and Allied Tables." Reprinted from the *Nautical Almanac* for 1937: H.M. Stationery Office.

Hayashi, K. 1926 "Sieben- und mehrstellige Tafeln der Kreis- und Hyperbelfunktionen." Berlin: Springer.

McDougall, J. and Stoner, E. C. 1938 *Philos. Trans.* A, **236**, 67.

Nordheim, L. 1934 "Müller-Pouillets Lehrbuch der Physik", **4**, 4, 277. Braunschweig.

Stoner, E. C. 1936*a* *Philos. Trans.* A, **235**, 165.

— 1936*b* *Phil. Mag.* **22**, 81.

— 1938*a* *Proc. Roy. Soc.* A, **165**, 372.

— 1938*b* *Proc. Leeds Phil. Soc.* **3**, 457.

— 1938*c* *Phil. Mag.* **25**, 899.

The nature of sliding and the analysis of friction

BY F. P. BOWDEN AND L. LEBEN

Laboratory of Physical Chemistry, Cambridge

(*Communicated by G. I. Taylor, F.R.S.—Received* 8 *August* 1938)

[Plates 17-24]

Although numerous measurements of kinetic friction have been made and some empirical laws have been formulated we have, as yet, no clear understanding of the mechanism of friction of sliding solids. This paper describes an attempt to investigate in more detail the physical processes that occur during sliding and the nature of the frictional force that opposes the motion. This work was done with unlubricated metallic surfaces. That is to say, no foreign films were artificially added; the surfaces were as clean as they could conveniently be prepared in the laboratory.*

Most of the previous methods used for measuring kinetic friction record only the average force necessary to maintain sliding. It was considered probable on theoretical grounds and from the results of some earlier more primitive measurements that the frictional force might not be constant. An apparatus was therefore constructed which was capable of measuring any fluctuations if they occurred. Experiments with dissimilar metals sliding together show that as a general rule the frictional force is indeed not constant but undergoes most rapid and violent fluctuations. The sliding process is not a continuous one; the motion proceeds by jerks. The metallic surfaces "stick" together until, as a result of the gradually increasing pull, there is a sudden break with a consequent very rapid "slip". The surfaces stick again and the process is repeated indefinitely. When the surfaces are of the same metal the behaviour is somewhat different. Large fluctuations in the friction still occur but they are comparatively slow and very irregular. The average value of the frictional force is considerably higher than that found for dissimilar metals and a well-marked and characteristic track is formed during the sliding.

---

* The behaviour of really clean metallic surfaces from which all trace of adsorbed films and gases is removed in a high vacuum will be described in a later paper. (See Bowden and Hughes, *Nature*, **142**, 1039.)

Simultaneous measurements of the surface temperature are also made. Earlier work (Bowden and Ridler 1935) has shown that the surface temperature of sliding metals may be very high. An analysis of this surface temperature shows that it, too, is changing rapidly during sliding and that there is a well-defined correlation between these changes and the frictional fluctuations. Every time a slip occurs there is a sudden rise in the temperature followed by a very sudden fall. The whole temperature "flash" may be over in less than a thousandth of a second.

It will be clear from these results and from the observations described in the next paper that metals do not slide uniformly. An intermittent clutching and breaking away of the metal surfaces can occur; the friction, the surface temperature, and the area of contact all vary rapidly.

Even if the surfaces are lubricated with mineral oils or other lubricants the behaviour may be essentially the same. The friction fluctuates and the movement again proceeds by a process of "stick and slip". This is consistent with the earlier observations that metal to metal contact can occur even under conditions of good boundary lubrication.

These observations have an important bearing on the theory of friction and lubrication. The detailed analysis of the frictional force shows that the "classical laws" of friction can be regarded only as crude approximations. Earlier theories have assumed that sliding is continuous and that only the surface atoms of the solid are involved. It is clear that neither of these assumptions is correct. The frictional effects are not confined to the surface. The sliding causes a permanent deformation of the solid which extends to a great depth beneath the surface so that the friction must also be dependent on the bulk properties of the solids. Any theory which involves the surface forces only must be incomplete, and we cannot expect such theoretical calculations to agree with experiment.

It is not possible at this stage to give a complete and quantitative theory of kinetic friction, but the detailed analysis of the fluctuating friction, surface temperature and area of contact, which is described in this and the subsequent paper, enables us to form some picture of the physical processes that occur during sliding. The work suggests that the kinetic friction between moving metals is due to a local welding together and a formation of metallic junctions between the metals. Since the size of these junctions is great compared with the dimensions of a molecule, their subsequent breaking causes a tearing and deformation of the metal which extends to a considerable depth. The exact behaviour depends upon the nature of the metals, and there is evidence that three distinct types of metallic junctions may be formed.

EXPERIMENTAL*

If fluctuations in friction are to be recorded it is essential that the friction-measuring device should have a high natural frequency and that the system which applies the load should have a natural frequency which is as high or higher. The apparatus which fulfils these conditions is illustrated diagrammatically in fig. 1. This diagram is roughly to scale.

The friction was measured between two surfaces, the lower of which was driven steadily on by a force much greater than the opposing frictional force, while the upper was connected to a fixed friction-measuring device. The lower surface *P* was plane and was screwed on to a base *R* which moved between parallel rails in the direction shown by the arrow in fig. 1. The rails were mounted on a thick insulating sheet of ebonite *E* which was screwed to a heavy steel block *F*. The surface was driven at a uniform rate by water pressure. The water was held under a pressure of five atmospheres in a 200 l. tank and a suitable system of needle valves enabled water to flow at any desired speed into the cylinders. The bore of these cylinders, which were obtained from an old aero engine, was 5 in. The incoming water drove the pistons forward at a uniform rate and the motion was transmitted to the lower surface. In this way a powerful uniform motion free from vibration was obtained. The influence of external vibrations was further minimized by mounting the whole apparatus on solid concrete.

---

* This section and part of the next revised 29 October 1938.

The upper surface was in the form of a small spherical slider *Q* shown in fig. 1. It was attached to a rigid duralumin arm *YZ* carried on a bifilar suspension *XX* of tightly stretched piano wire. The load was applied to *Q* in a direction normal to the surface *P* by means of the circular spring *R*. The latter consisted of a piece of watch spring bent round to a circle as shown in the diagram. The extent of compression of this spring in the vertical plane determined the load and this could be maintained constant by locking the nuts *L*, *L'*. A flat horizontal piece of watch spring *T* was fixed between the arm *YZ* and the bottom of the circular spring *R*, in order to keep *Q* in the plane of *YZ*. The slider *Q*, which was threaded on its upper part, passed through holes in the springs *R* and *T* and was rigidly held in place by a nut.

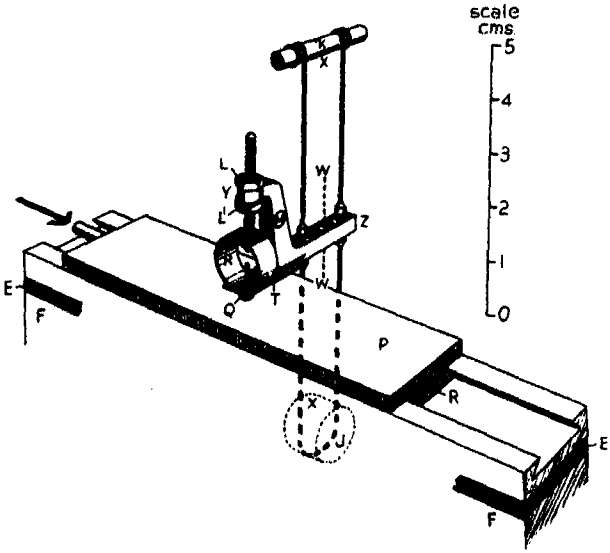


FIG. 1.

The bifilar system *XX* was carried on a rigid frame. The bifilar suspension consisted of a length of piano wire which passed over a pulley *J* and which was tightened by means of the axle *K*. The instrument was damped by pads of piano damping felt pressed against the wires *XX*. Electromagnetic and oil damping had been tried previously, but this method was found to be best for the purpose. The whole frame was mounted so that it could be fixed at any distance from the block *F*. This allowed a number of different runs to be conducted on parallel tracks on the same lower surface.

The load was applied as already described. The exact value was determined by applying a vertical pull to Q using a suitable hooking device screwed on to the upper part of the slider. The force needed to break contact between the surfaces (determined electrically) was equal to the normal load.

Movement of the lower surface P under the action of the hydraulic pressure gave rise to frictional forces between P and Q. This caused the arm YZ to be deflected about the vertical axis WW, and the extent of deflexion is a measure of the tangential force between the surfaces, i.e. of the friction. This deflexion was measured by means of a beam of light reflected from a small mirror M. The reflected beam fell on to a horizontal opening in a moving-film camera. A horizontal cylindrical lens in this opening focused the beam on to the 15 cm. wide oscillograph paper used for recording. The light path was 6·5 m. long, so that the apparatus was very sensitive. The friction-measuring device was calibrated directly by removing the lower surface, applying known horizontal forces to Q in the direction of motion, and measuring the corresponding deflexions of the spot of light on the film. A cathetometer was focused on Q, and the actual movement of Q under the action of the applied horizontal forces also measured. The relation connecting the distance moved by Q and the corresponding deflexion of the spot of light on the film was then known. The photograph thus obtained gave a continuous record of the tangential forces which were acting between the surfaces. This tangential force divided by the normal load is represented on the photographs as the coefficient of friction.

Time marks of 1/64th sec. (i.e. 0·016 sec.) were thrown on the edge of the camera film by a stroboscopic disk and a heavier mark every 30/64th sec. (i.e. 0·47 sec.). In the reproduced photographs the individual time marks are not always clearly visible, and each thirtieth division has been accentuated. In all the photographs the time interval between successive marks corresponds to 0·47 sec., except in figs. 7*a* and 7*b*, where they occur every 0·1 sec.

The friction-measuring device incorporates two vibratory systems. One measures the tangential reaction between the surfaces and its frequency is controlled by the bifilar suspension XX. The other applies the normal force between the two surfaces and is controlled by the circular spring R. The flat spring T does not affect the load since its position remains constant, but its rigidity tends to increase the natural frequency of this system. By the nature of this apparatus these two systems are independent of one another except in so far as the latter system will contribute towards the inertia of the former. Nevertheless, the masses of the parts have been kept sufficiently low for both systems to have a high frequency of the order of 1000. This was tested in the completed apparatus by allowing the arm YZ to vibrate with the slider Q off the surface, and recording the movements in the camera. The trace showed that the arm was vibrating with a frequency of the required order. The frequency of the system which applied the load was still higher. The damping was so arranged that the vibrations died out very rapidly. It is clear that a system of this type is capable of recording any changes in the tangential force between the surfaces provided that the frequency of these changes is not greater than the natural frequency of the measuring system.

Measurement of surface temperature

The surface temperature was measured by using the rubbing contact of two different metals as a thermocouple, and the method was essentially the same as that previously used (Bowden and Ridler 1936). When the slider Q is a different metal from the sliding surface P, any fluctuations in surface temperature during sliding will give rise to corresponding thermoelectric fluctuations. These impulses were recorded simultaneously with the frictional fluctuations, using an Einthoven galvanometer adjusted to its highest natural frequency. Since this is also its position of minimum sensitivity the impulses had first to be magnified and a simple two-stage amplifier was used for this purpose.

Before interpreting the results of the experiments it is necessary to consider two possible sources of error. First, small changes of resistance take place between the surfaces during sliding and this might alter the apparent e.m.f. The contact resistance was only a few hundredths of an ohm. Since the thermocouple was always in series with a high resistance, any changes in the contact resistance between the surfaces could have no appreciable effect. A second possible source of error is connected with the fact that the thermoelectric e.m.f. varies with the state of working of the metals which form the couple. Because of this, any "working" of the metal during sliding will cause a change in their thermoelectric power. It is clear, however, that this effect is also very small. The thermoelectric force between worked and annealed specimens of a metal varies slightly with the direction of measurement on the specimen, but it is always small. For constantan it is about $0{\cdot}7 \times 10^{-6}$ V/° C, and for iron and steel it is of the same order (Elam 1935). The average thermoelectric e.m.f. of the constantan steel couple formed by the surfaces was about 25×10^{-6} V/° C, so that the state of working of the surfaces can only affect the real thermoelectric e.m.f. by a few per cent. This variation is within the limits of experimental error. Bridgman (1918) has shown that pressure can also cause a change in the thermoelectric power of a metal, but the magnitude of this effect is even smaller than that discussed above.

Preparation of surfaces

The lower surface was prepared by grinding on flat lead laps with various powders using the technique described in earlier papers. The upper curved slider was turned in a lathe and polished. The surfaces were then cleaned with rouge and water, washed in caustic soda, rewashed with water, and finally cleaned with specially purified acetone. If water were added to these surfaces it would spread as a thin continuous film (see Hardy).

The steel used for most of these experiments was specially prepared for us by Dr W. H. Hatfield. Its composition is: carbon 0·13 %; silicon 0·21 %; manganese 0·59 %; sulphur 0·018 %; phosphorus 0·015 %; chromium 0·05 %; nickel 3·42 %. The specimen was taken from the bottom of the ingot and is very uniform in texture.

RESULTS

The friction of steel on steel

Fig. 2 (Plate 17) shows a typical trace taken with the lower surface moving at a constant speed of 0·006 cm./sec., and with a load of 400 g. between the surfaces. The original trace was 15 cm. in width, and this has been suitably reduced in the photograph. The passage of time is from right to left of the trace. The deflexion of the arm YZ (i.e. the tangential force between the surfaces) is recorded at right angles to this direction. This tangential force divided by the normal load gives the coefficient of friction, and the value of this is marked on the ordinates. The dark trace therefore represents the variations of friction with time. It will be seen that the force opposing motion is not constant but is fluctuating violently during sliding. These fluctuations all consist of a period AB during which the tangential force is increasing steadily, followed by a very rapid fall BC. A subsidiary experiment had shown that a movement of 0·0025 cm. of the slider Q caused a deflexion of 1 cm. of the light spot on the original film, so that the optical magnification was 400. The slope of the line AB was measured with reference to the time axis, and calculation showed that the top surface Q is moving forward with a velocity of 0·006 cm./sec. This velocity is identical with that of the lower surface, and it is clear that the two surfaces are "sticking" together. The upper surface is being carried forward by the lower one, and no relative motion is occurring. Suddenly, at the point B, a rapid slip of the upper surface Q occurs. When the frictional force has fallen to C the surfaces "stick" together again. The process is then repeated indefinitely.

The speed of the camera is not sufficient to allow of accurate recording and analysis of the slip BC. On the traces it appears as a faint line, and its duration can be estimated to be of the order of 1/1000th sec. It is important to note that this is approximately the same as the free period of the instrument. In fact, the natural frequency of the bifilar system sets a limit to the rapidity with which the slip occurs. It is impossible for the arm YZ and the top surface Q to move faster than the suspension XX will allow. A cathode-ray method is necessary for a more detailed analysis of these very rapid changes.

Fig. 3 (Plate 17) shows a trace taken with the lower surface moving at a higher speed of 0·3 cm./sec. The load was again kept constant at 400 g. In this experiment the camera was run at a slightly greater speed in order to facilitate the analysis of the traces. The friction is fluctuating violently, and measurement again shows that the motion is proceeding by a process of "stick-slip". Again AB corresponds to the "stick" and BC to the "slip". Careful measurement of the slope of the portion AB shows that during this period the speed of motion of Q was now 0·3 cm./sec. and so corresponds to that of the lower surface. Hence no relative motion is occurring. Owing to the increased speed of the lower surface the "stick-slips" now occur much more frequently on the photograph. It is also interesting to note that the *average* value of the friction varies in an irregular fashion over a range of about 20 % during the course of the run.

The actual length of the slip between the surfaces can be determined by measurement of the corresponding length BC on the original photograph since the optical magnification of the system is known. In the case of steel on steel an average slip was 0·003 cm.

These traces were obtained on a flat surface finely ground with 320 carborundum. In other experiments the surfaces were highly polished so that a typical Beilby layer was formed. Again the fluctuations in friction occurred and the behaviour was essentially the same.

The friction of different metals on steel

A series of different metallic sliders were prepared and the experiment repeated using each slider in turn on the steel plate. In each case a load of about 400 g. was used and approximately the same speed (0·003–

0·006 cm./sec.) employed. The exact speed in each case determines the slope of the dark friction trace in the corresponding photograph. The sliders were made of the following metals:

Metal	Melting-point °C	Vickers hardness
Molybdenum	2470	200
Nickel	1452	85
Steel	—	166
Constantan	1290	113
Silver	960	32
Aluminium	658	14
Zinc	419	34
Lead	327	4·0
Tin	232	6·0
Bearing metal	180	20
Wood's alloy	70	16

The results obtained are shown in fig. 5 *a–j* (Plates 18, 19). It will be seen that "stick-slips" occur in each case, and the magnitude of the slip increases regularly as the melting-point of the upper slider falls. Figs. 3 and 4 (Plate 17) show that this holds true even if the speed is increased by a factor of a hundred.

The track caused by sliding

We should expect that the wear and tear of the metal surfaces during sliding should bear some relation to the frictional behaviour. A microscopic examination was therefore made of the track formed on the steel plate by the sliders of different metals. Fig. 6 *a–h* (Plate 20) shows photomicrographs of the various scratches, all taken at a magnification of 70. The tracks are indicated by arrows; when two arrows are shown the track lies between them. The sliding metal has passed over this surface once only. These photographs correspond to friction traces given in fig. 5 and were taken during the same runs.

The simultaneous measurement of friction and surface temperature

Any discontinuity in the sliding between two metals should lead to a similar discontinuity in the evolution of heat and hence to fluctuations in the surface temperature. An experiment with constantan sliding on steel showed that this is indeed the case. Fig. 7*a* (Plate 21) is a trace showing the simultaneous variations of friction and surface temperature. Analysis of the curves shows that every time a slip occurs there is a sudden rise and fall in the surface temperature. This is shown more clearly in fig. 7*b* (Plate 21), obtained with the surfaces moving at a comparatively slow speed. A single rise and fall in temperature is seen coincident with the slip between the surfaces.

The experiment has been repeated with Wood's alloy, tin, and silver sliding on steel. The results are essentially similar to those obtained with constantan.

Other combinations of metals

In the experiments so far described the lower surface was always of steel. Further experiments have been carried out using other metals. Measurements were made with Wood's alloy, tin, silver, steel, platinum, and molybdenum, on plates of tin, silver, copper, nickel, platinum, and molybdenum. The results are essentially similar to those obtained on steel, and it can be stated that in general the sliding of two dissimilar metals is discontinuous, the motion proceeding by a process of stick and slip.

The sliding of similar metals

When *similar* metals are slid on one another the results are different. The frictional behaviour of silver, copper, nickel, platinum, and molybdenum on themselves are shown in figs. 8 and 9 (Plates 22, 23); the corresponding photomicrographs of the tracks are shown in fig. 10*a–e* (Plate 24). Essentially the same behaviour was also observed with tin on tin.

The friction of flat surfaces

All the experiments so far described were carried out with a small round slider moving on a flat surface. It is important to see whether these observations apply to the more general case of flat surfaces rubbing together, and experiments were carried out to test this. The upper surface in this case was in the form of a flat disk about 1 cm. in diameter and 1 mm. thick. The lower face was ground flat. The disk was attached to the arm *YZ* in a manner similar to that used for the curved slider.

Experiments with a steel flat on the steel plate showed that the sliding again proceeded by a process of "stick and slip"; the magnitude of the fluctuations was essentially unaltered. Subsequent examination of the lower steel plate showed several parallel scratches. The two surfaces were obviously touching at several points at any instant of motion. Nevertheless, the nature of the sliding remained the same.

The experiment was repeated for the case of a tin flat sliding on the steel surface. The motion again proceeded by a process of "stick and slip". The slips were large and the behaviour was essentially the same as that observed with a curved slider of tin. A wide smear of tin was formed on the steel. Microscopic examination of this track showed however that contact actually occurred only at many small spots distributed over the entire area of apparent contact.

The friction of lubricated surfaces

It is to be expected that the presence of a lubricant will affect the nature of the sliding between two bodies. A preliminary investigation of this problem has been made employing the apparatus described above and using a steel slider on a steel surface. With mineral oils it is found that the motion still proceeds by a series of "stick-slips", and the general behaviour is similar to that of unlubricated surfaces except that the length of the slips is different. With certain molecules such as the long-chain fatty acids the "stick-slips" disappear completely, sliding becomes continuous, and the kinetic friction shows no fluctuations.

DISCUSSION

These experiments show clearly that the frictional force between various metals sliding on steel is not constant but is fluctuating violently. The motion is not continuous, but proceeds by a process of "stick and slip". It might be considered that this is due to small scratches and irregularities on the surface, but the evidence is strongly against this. A careful measurement and a microscopic examination of the surface showed that there was no correlation between the discontinuities of motion and the surface scratches. Experiment also showed that "stick-slips" occurred, and the behaviour was essentially the same on highly polished surfaces which were free from scratches and covered with the typical Beilby layer. The fact that the magnitude of the stick-slips depends upon the metals used, although the lower steel plate was ground in the same way in every case, lends additional support to this view. We should expect no difference in the slip if this discontinuity were a result of mechanical interlocking of the surface asperities. Finally, the work on lubricants has shown that with certain molecules stick-slips disappear altogether, although the thickness of the applied film is much less than the dimensions of the asperities. The work suggests that these discontinuities in motion are inherent in the sliding of metals. The kinetic friction may, with important reservations, be regarded as a series of static frictions. An examination of fig. 5*a*–*j* (Plates 18, 19) for various metals on steel shows that the extent of the rapid slip is governed by the melting-point of the metal. The lower the melting-point the larger the slip. There is of course an approximate correlation between the various mechanical properties of metals and their melting-points, so that we might expect a systematic relationship similar to that found above. The values for the Vickers hardness given in the table are those obtained by direct measurement on each specimen. It is of course possible that this value may be modified by the surface working. It is clear that the hardness of the metals must play some part, but the results suggest that the melting-point is of even greater significance. For example, Wood's alloy is harder than tin yet it gives a *larger* slip because the melting-point is lower. In the same way, zinc, which is harder than aluminium but has a lower melting-point, gives a larger slip.

An examination of the photomicrographs of the track is interesting (fig. 6*a–h*, Plate 20). It will be seen that the highest melting metals, molybdenum to silver, all plough out a groove. The width of this groove is greatest for the metal of highest melting-point. The behaviour of the lower melting metals, zinc to Wood's alloy, is however quite different. These metals do not scratch the surface, but are themselves wiped off on to it to form a well-defined smear. The transition from *type 1* to *type 2* occurs with aluminium which is intermediate in behaviour, sometimes scratching the steel and sometimes being smeared on to it. With other combinations of metals the general behaviour is the same. Sliding is again discontinuous and the extent of the rapid slip is dependent on the melting-point of the solids. The lower melting metals give the greater slips, and again two distinct types of track are observed.

When the two surfaces are of the same metal (*type 3*) the behaviour is different from either of the above cases. The nature of the frictional force between the moving surfaces of silver, copper, nickel, platinum, and molybdenum, is shown in figs. 8 and 9. Again the frictional force is not constant; very large fluctuations occur, but the nature of these fluctuations is quite different. A very sudden rapid slip no longer occurs, but the fluctuations are much slower and very irregular. This characteristic irregularity is particularly well marked in the photographs taken with the surfaces moving comparatively fast (compare with figs. 3 and 4 showing fast runs of steel and lead on steel). It will also be observed that the average value of the friction, in so far as the average value of these fluctuations can be made, is considerably greater than that found for dissimilar metals.

The nature of the track formed is shown in fig. 10*a–e*. It is clear that this track is quite different from either of the two tracks formed by dissimilar metals. A groove is formed in the flat surface, but it looks very different from the groove cut by a metal of higher melting-point. The general appearance would suggest that the metal within the groove has flowed during sliding and then solidified. This third type of behaviour is characteristic of similar metals whatever their hardness or melting-point. Steel is an apparent exception. With steel on steel the behaviour corresponds to *type 1*, i.e. rapid slips and torn track. This exception is probably more apparent than real, since the surface of this alloy is not homogeneous and over any small area the two rubbing surfaces are not identical.*

Although oxide films must be present in varying thicknesses on all these metal surfaces and must play some part they do not appear to govern the type of sliding. On platinum, for example, the oxide film is very thin, on silver it is easily broken down, on tin and molybdenum it must be quite thick, yet the results for both similar and dissimilar metals show that this factor does not govern the type of sliding.

The experiments also show that the surface temperature is fluctuating continuously during sliding. Analysis shows a definite correlation with the frictional changes; every time a slip occurs there is a very sudden rise followed by a very rapid fall in the temperature. The whole temperature "flash" may be over in less than one-thousandth of a second. The apparent temperature rise is small, but two factors prevent us from evaluating the real surface temperature. The temperature flash is so rapid that it cannot be followed by the Einthoven galvanometer. Also the actual points of contact will not all be at the same temperature. Short-circuiting will occur through the colder points and the e.m.f. recorded will only be a mean value.

Most of these measurements were made with a curved surface sliding on a flat plate. Theoretical considerations lead us to expect that the phenomena should not be very different when two flat surfaces slide together. This is confirmed by the experiments that have been carried out with tin and steel flats on steel. Although these two metals give different types of sliding the results in each case were essentially the same as those observed with curved surfaces.

It is clear that the moving parts of these systems are not symmetrical since the restraint is not the same for each surface. The upper surface is connected to a system possessing low inertia, high frequency, and an appreciable degree of elastic motion. The lower surface, on the other hand, has a large inertia and a negligible freedom of motion. In these experiments, therefore, the two surfaces cannot be treated as equivalent to one another, and we would expect that the order in which they are slid on one another will have some bearing on the nature of the sliding. The extensive experiments with rounded sliders on flat surfaces have shown that this is indeed the case; the extent of the fluctuations in friction between two metals is often altered by reversing the order of sliding. Similar considerations must apply in some engineering problems, e.g. when a moving system such as a journal slides in a fixed bearing.

---

* [*Note added in proof* 16 December 1938.] This is confirmed by experiments on two similar surfaces of an austenitic chromium steel of uniform structure. In this case the friction and wear were of *type 3*.

THE MECHANISM OF KINETIC FRICTION

In 1699 Amontons published the result of his experimental investigation of the friction of unlubricated solids. He found that the frictional force was independent of the area of the surfaces and was directly proportional to the normal load. In fact he concluded that the frictional force was always equal to one-third of the normal load. To explain this result he assumed that irregularities on the surfaces of the two bodies interlocked, and the relative motion necessitated lifting the load from one position of interlocking to another. This caused a loss of energy which was manifested as a frictional force. Amontons's conclusions were tested by many investigators, notably De la Hire (1732) and Euler (1750) The latter agreed with Amontons in giving to all surfaces a frictional coefficient of one-third. The most systematic work was done by Coulomb (1785) who examined the influence of a large number of variables on the friction. He agreed that the friction was proportional to the load and made the new observation that it was independent of the velocity of sliding. Further work was done by Rennie (1829), Morin (1835), and Hirn (1854). Hirn was the first to distinguish carefully between lubricated and unlubricated solids, and he observed that the effect of velocity, surface area and load differed in the two cases. Few precautions were taken by any of these experimenters to obtain clean and reproducible surfaces. For example, Rennie's surfaces were almost certainly contaminated, since the value he obtained for the coefficient of friction was very low (*ca.* 0·1 or less). All these earlier workers agreed with Amontons in assuming that friction was caused by the interlocking of asperities. Rennie pointed out that a more general theory should take into account the bending and fracture of these asperities. The action of lubricants was explained by assuming that the lubricants filled these irregularities and at the same time, by some means, made them more "slippery".

These earlier theories were questioned when Lord Rayleigh suggested that the difference between a polished surface and the surface of a fluid

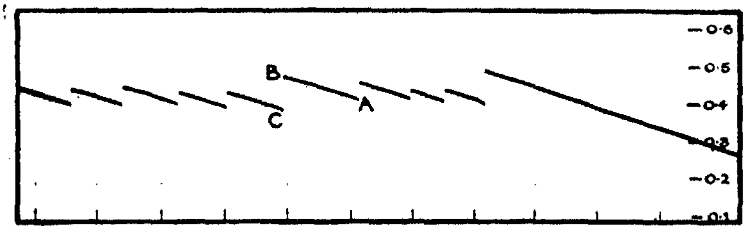


FIG. 2. STEEL ON STEEL. SLOW.

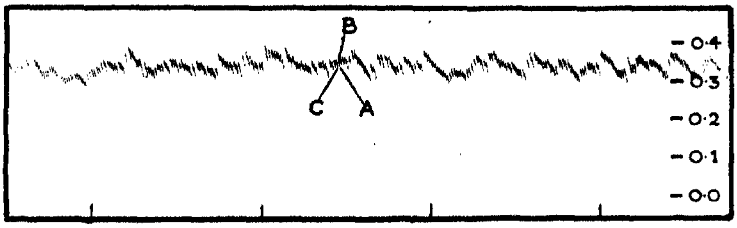


FIG. 3. STEEL ON STEEL. FAST.

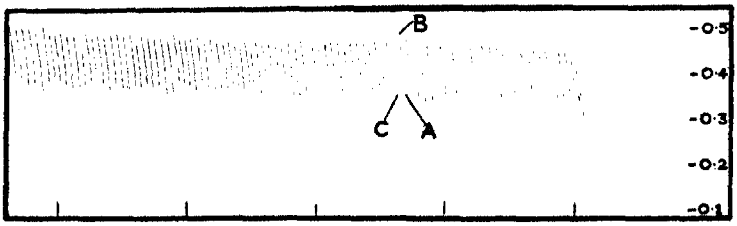


FIG. 4. LEAD ON STEEL. FAST.

← TIME : ONE DIVISION = 0·47 SEC.

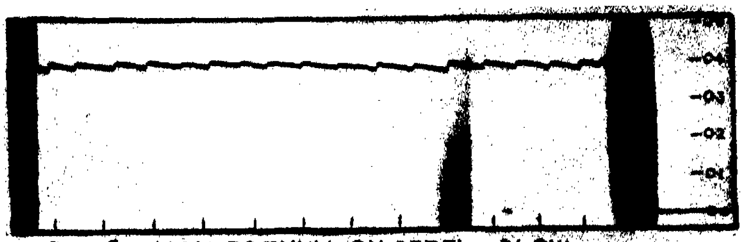


FIG. 5a. MOLYBDENUM ON STEEL. SLOW.

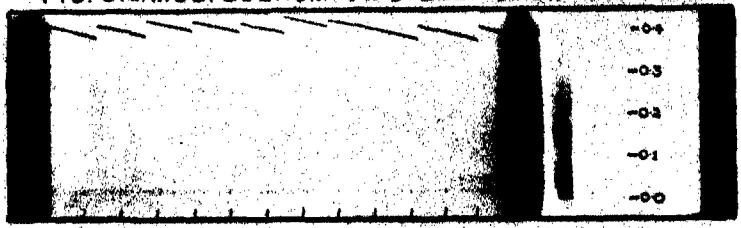


FIG. 5b. NICKEL ON STEEL. SLOW.

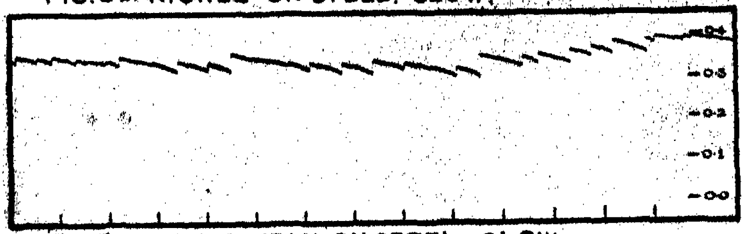


FIG. 5c. CONSTANTAN ON STEEL. SLOW.

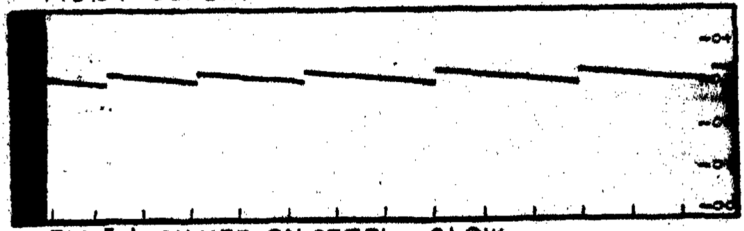


FIG. 5d. SILVER ON STEEL. SLOW.

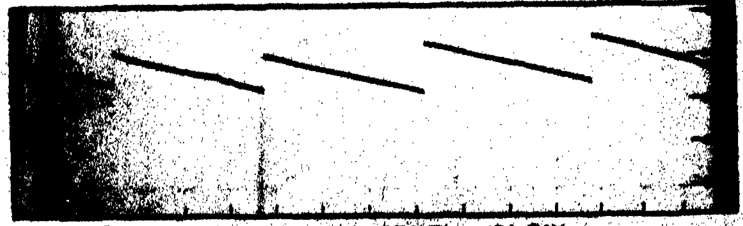


FIG. 5e. ALUMINIUM ON STEEL. SLOW.

COEFFICIENT OF FRICTION.

← TIME: 1 DIV. = 0.47 SEC.

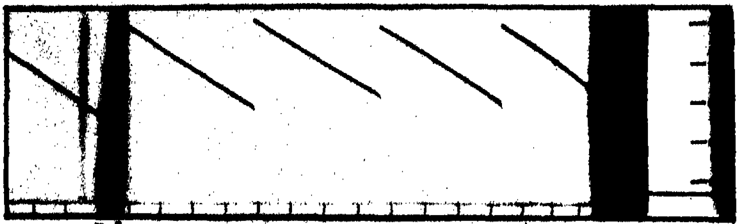


FIG. 5f. ZINC ON STEEL. SLOW.

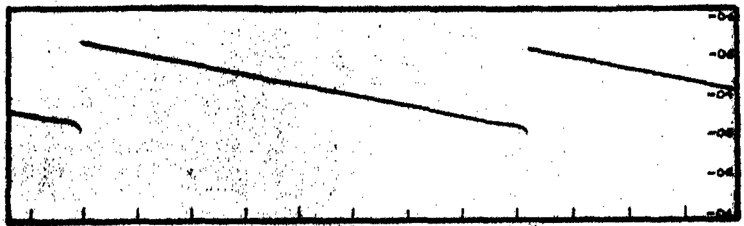


FIG. 5g. LEAD ON STEEL. SLOW.

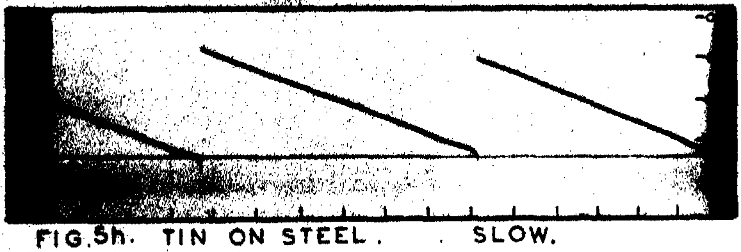


FIG. 5h. TIN ON STEEL. SLOW.

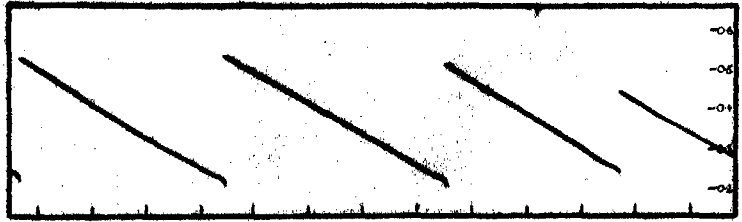


FIG. 5i. BEARING METAL ON STEEL. SLOW.

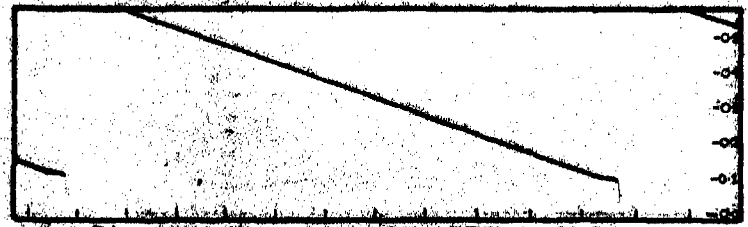


FIG. 5j. WOODS ALLOY ON STEEL. SLOW.

COEFFICIENT OF FRICTION.

← TIME: 1 LARGE DIV. = 0.47 SEC.; 1 SMALL DIV. = 0.016 SEC.

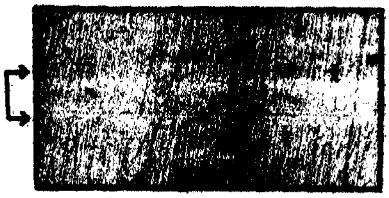


FIG. 6a. MOLYBDENUM
ON STEEL.

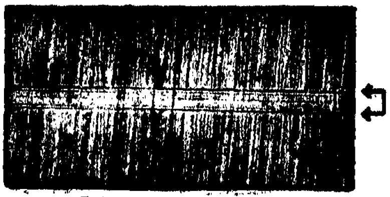


FIG. 6b. NICKEL
ON STEEL.

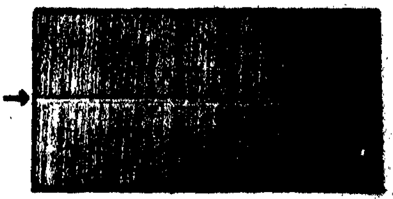


FIG. 6c. STEEL
ON STEEL.

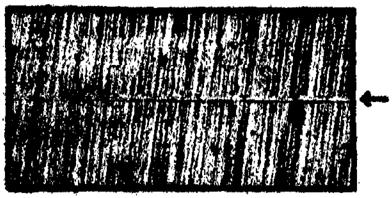


FIG. 6d. CONSTANTAN
ON STEEL.

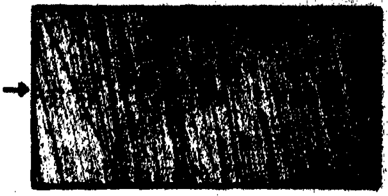


FIG. 6e. SILVER
ON STEEL.

FIG. 6f. ALUMINIUM
ON STEEL. (1)

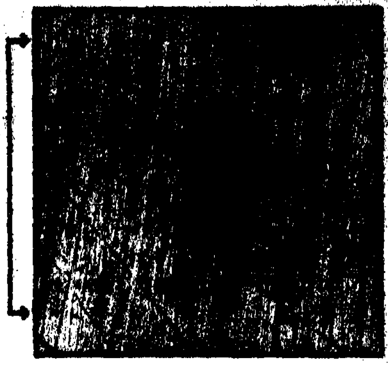


FIG. 6g. ALUMINIUM
ON STEEL. (2)

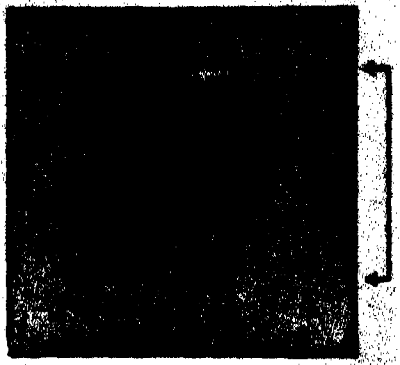


FIG. 6h. TIN
ON STEEL.

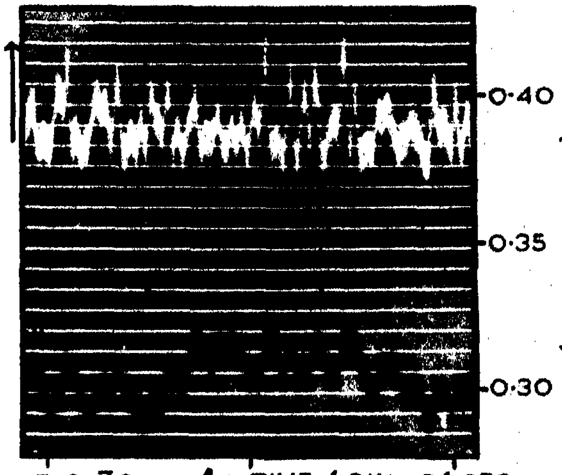


FIG. 7a. ← TIME: 1 DIV. = 0.1 SEC.

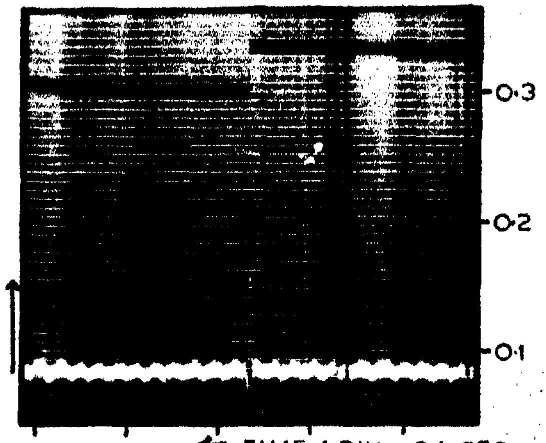


FIG. 7b. ← TIME: 1 DIV. = 0·1 SEC.

CONSTANTAN ON STEEL.

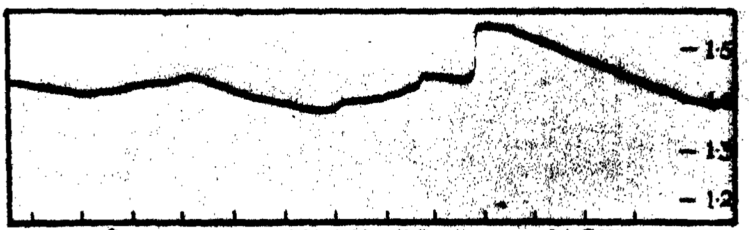


FIG. 8a. SILVER ON SILVER. SLOW.

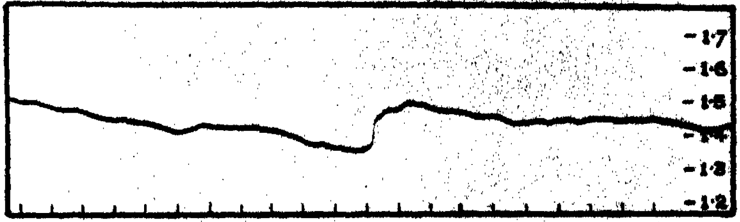


FIG. 8b. COPPER ON COPPER. SLOW.

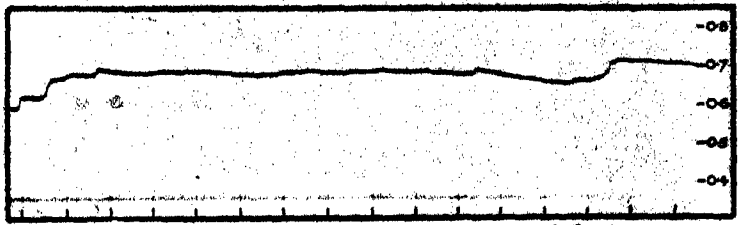


FIG. 8c. NICKEL ON NICKEL. SLOW.

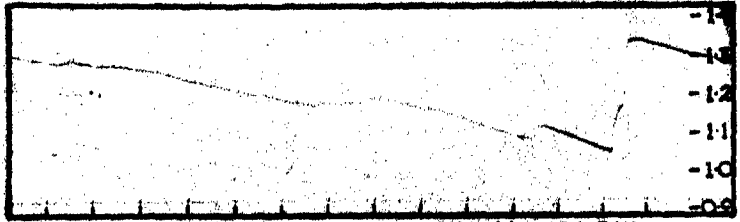


FIG. 8d. PLATINUM ON PLATINUM. SLOW.

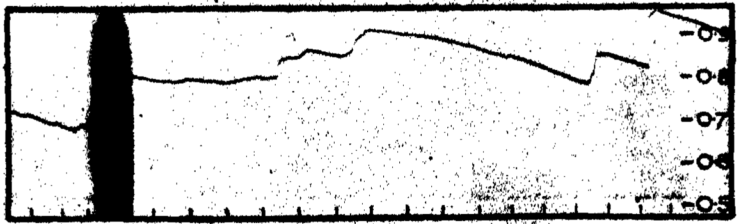


FIG. 8e. MOLYBDENUM ON MOLYBDENUM. SLOW.

← TIME: 1 LARGE DIV.=0.47 SEC., 1 SMALL DIV.=0.016 SEC.

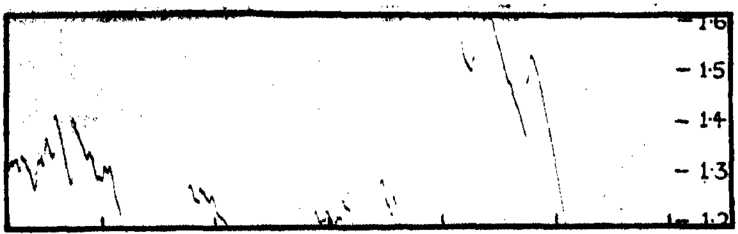


FIG. 9a. SILVER ON SILVER. FAST.

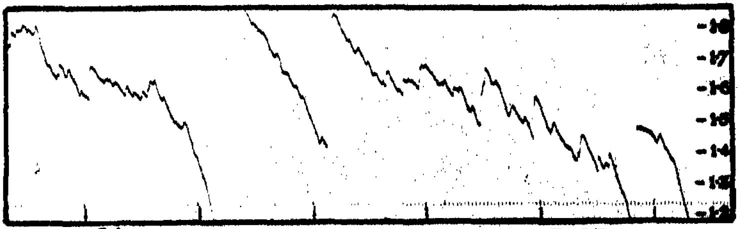


FIG. 9b. COPPER ON COPPER. FAST.

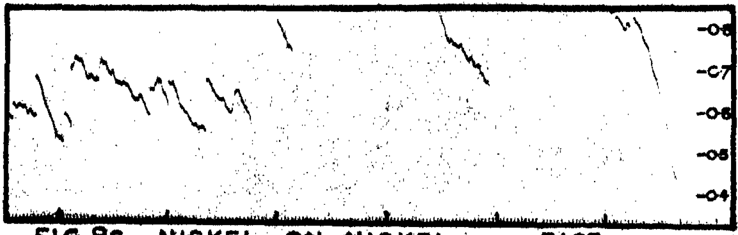


FIG. 9c. NICKEL ON NICKEL. FAST.

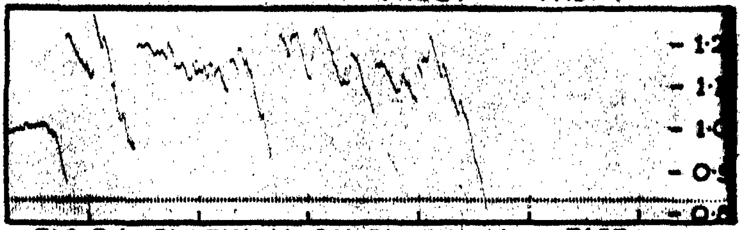


FIG. 9d. PLATINUM ON PLATINUM. FAST.

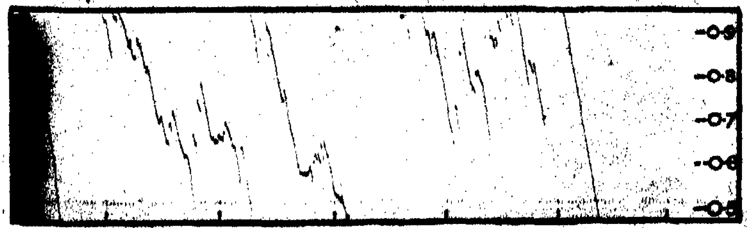


FIG. 9e. MOLYBDENUM ON MOLYBDENUM. FAST.

← TIME: 1 LARGE DIV. = 0.47 SEC.; 1 SMALL DIV. = 0.016 SEC.

COEFFICIENT OF FRICTION.

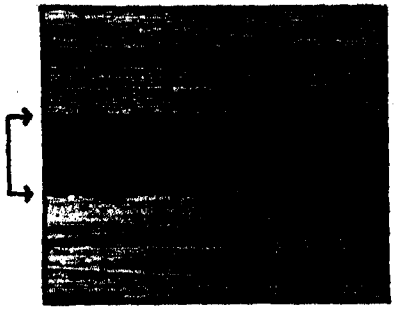


FIG. 10a. SILVER ON SILVER.

FIG. 10b. COPPER ON COPPER.

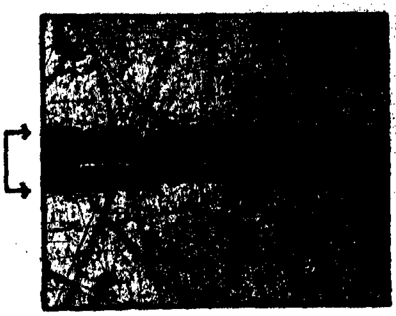


FIG. 10c. NICKEL ON NICKEL.

FIG. 10d. PLATINUM ON PLATINUM.

FIG. 10e. MOLYBDENUM ON MOLYBDENUM.

might not be very great. This view was confirmed by Beilby who showed that polishing and grinding were two essentially different processes. The more recent view that friction had its origin in surface forces and was due to molecular cohesion between the solids was then introduced. Ewing (1892) considered that friction was due to the reaction of molecular forces following the molecular displacement. This theory has also been put forward and has received strong support from the extensive work of Sir William Hardy on static friction. He found that when two unlubricated glass surfaces slid together an actual tearing of both faces occurred. The scratched line was composed of an irregular collection of very shallow pits and very thin plates torn from the opposite face. Hardy concluded that the two surfaces seized under the action of the forces of cohesion so that tearing occurred during the subsequent motion. Unlubricated friction was therefore due primarily to the forces of cohesion between the surfaces. The results he obtained with lubricants were explained by assuming that the lubricant caused, in a very definite and quantitative way, the reduction of the molecular field of force at the surface of the solid.

An interesting attempt to correlate the interaction during sliding between the surface molecules at the surface of two unlubricated solids has been made by Tomlinson (1929). He considered that the friction was due to the energy dissipated when the molecules were forced into each others atomic fields and were then separated. He made the further assumption that the molecular fields of force were approximately the same for all substances, and the area of molecular contact could be calculated from Hertz's equation for elastic deformation. He found approximate agreement of experiment with theory, but many of his experimental values for friction were very different from those found by other workers. A molecular theory of friction has also been put forward by Derjaguin. He neglected intermolecular attractions and replaced the repulsive forces by quasi-rigid bonds.

The physical effects of sliding

These theories differ somewhat, but they all make the assumption that friction is due to the surface fields of force only. It was stated in an earlier paper (Beare and Bowden 1935) that the frictional effects were not confined to the surface layers of the solid but caused distortion and abrasion to considerable depth below it, that the physical processes which occurred during sliding were too complicated to yield easily to simple mathematical treatment, and that a more detailed knowledge was necessary before a quantitative calculation of friction could be made. An examination of the surface after the slider has passed over it shows that the friction is certainly not a surface effect only; the disturbance has penetrated to a depth of many thousands of atoms. For a complete interpretation of the friction between dry metals it is therefore necessary to take into account some of the properties of the metals in bulk. The experiments with different metals on steel confirm this point; they show that the nature of the sliding varies with the metal employed. The behaviour of any metal must be dependent on its bulk properties.

Moreover, it has usually been assumed that sliding is a continuous process; it is clear that this is not the case. An intermittent clutching and breaking away of the surfaces is occurring during sliding, and the frictional force, area of contact and local surface temperature, are all fluctuating violently. This surface temperature may easily become high enough to cause a softening and melting of the solid surfaces. Any complete theory of friction must take these effects into account. These remarks apply primarily to the sliding of unlubricated surfaces but there is evidence that even in the presence of lubricants metallic contact can occur through the film and that frictional and temperature fluctuations and abrasion may occur.

This more detailed analysis of friction and sliding shows that the process is a complicated one involving both the surface and the bulk properties of the moving solids. It is not possible at this stage to give anything like a quantitative or complete theory. Our knowledge of the detailed physical processes that occur during sliding is at present far too incomplete. The results of this and the following paper suggest that the friction between unlubricated surfaces is due primarily to the formation of local metallic junctions, the dimensions of which are probably large compared with the dimensions of a molecule. Relative motion necessitates the breaking of these junctions with consequent distortion and tearing of the metal to a considerable depth beneath the surface. These junctions are formed by a local welding together of the surfaces. Although the surfaces may be highly polished there will still be peaks and undulations present. Contact will occur only locally at the summit of these irregularities so that the area of intimate contact is very small. Even with lightly loaded surfaces the local pressure will be very great. This local pressure, under some conditions, may be sufficient to case a welding together, but it will be greatly assisted by the local high temperatures which are developed during sliding.

A more detailed examination of the track confirms the assumption that metal-metal contact may occur between the two moving surfaces. Thus for steel sliding on steel under a load of 400 g. the width of the scratch is 2×10^{-3} cm. Assuming the region of contact to be a circle with this length as diameter, we obtain a real contact area of about 3×10^{-6} sq. cm. and a mean pressure of about 130,000 kg./sq. cm. The stress at which plastic flow of a metal occurs will vary considerably with the state of working of the metal. For a carefully annealed mild steel Taylor gives a value of about 3000 kg./sq. cm. Other values quoted by Nadai for various steels vary from 4000 to 5000 kg./sq. cm. The existing pressure in the region of contact is therefore well able to cause the flow which must occur when a metal-metal junction is formed. More precise measurements of the real area of contact are given in the next paper.

Since a frictional force of 150 g. is being exerted across this surface of contact we find that a shear stress of 50,000 kg./sq. cm. is acting. This is much greater than the ordinary shear strength of steel (6000 kg./sq. cm.—roughly half the tensile strength). The experiments of Bridgman have shown, however, that under high confining pressures, of the order of 50,000 kg./sq. cm., the shear strength of materials rises and may even increase up to tenfold its normal value. In this connexion it is interesting to note that Bridgman found a well-marked welding on to the compressing pistons of the material under compression. Further, he found that under these conditions of high confining pressure in many cases plastic flow during shear was discontinuous, "chattering" occurring. In a recent paper Elam describes some tensile experiments in which steel test-pieces were extended very slowly. It was found that the extension was not continuous but that it took place in a series of jumps.

Three types of sliding

The experiments described in this paper suggest that the local metallic junctions may be of three types. The first type is characteristic of a slider of a hard high-melting metal sliding on a softer surface of lower melting-point. The motion proceeds in jerks, but the rapid slip is comparatively small. The surface of the high-melting metallic slider is not perceptibly changed, but a well-defined track is torn out of the lower melting metal. The next paper describes the behaviour of the conductance during sliding and this is very characteristic. It will be shown that when a rapid slip occurs, the conductance, and hence the area of contact, *falls suddenly to a minimum value*. During the stick, as the pull on the surface increases, the area of contact *increases* until the sudden slip again occurs when it again falls rapidly. The process is then repeated. The result suggests that surface irregularities of the hard high-melting metal penetrate the softer one, and the resistance to motion is due mainly to a ploughing out of the soft metal. As a crude analogue we may compare this to the pulling of a ploughshare or a number of ploughshares through hard soil. As the pull on the ploughshare increases its surface is pulled into closer contact with the earth and the area of intimate contact increases. Suddenly the earth yields and the ploughshare jerks forward. The strain is momentarily diminished and the pull and the area of contact fall to a lower value. The process is then repeated.

In *type 2*, which is characteristic of a low-melting slider moving on a metal surface of high melting-point, the behaviour is quite different. The motion again proceeds in jerks, but the extent of the rapid slip is very much larger (in the case of Wood's alloy—m.p. 70° C—on steel it is some fifty times as great as that of molybdenum—m.p. 2470° C—on steel). The flat surface of the high-melting metal is not torn or abraded. It is apparently undeformed but obvious blobs and smears of the low-melting metal are wiped on to it. Conductance measurements (see next paper) show that the area of contact is a *maximum the instant after the slip has occurred.* During the stick, as the pull increases, the area of contact is now observed to *decrease* until the sudden slip occurs again, when it again rises to a maximum. We may again make a crude picture and suggest that the high pressure aided by the heat liberated during the rapid slip solders or welds the low-melting metal on to the other. The extent of the metallic contact and hence the electrical conductance is then at a *maximum.* As the pull increases the soldered junctions are drawn out thinner and thinner, the conductance decreasing, until the junctions break. Slip occurs again and the process is repeated. The break will naturally occur at the soldered joints; the high-melting metal will be left untouched with the blobs of the solder adhering to it.

The behaviour of surfaces of the same metal (*type 3*) is however very different. Here the friction is very much higher and large fluctuations occur in it. No rapid slips are, however, observed. A large track is torn out of the flat surface, the metal within this groove having a very torn and distorted appearance. The behaviour of the electrical conductance is also characteristic. The general value of the conductance is higher, showing the area of contact is greater, and no rapid changes are observed. It is clear that in this case local high pressure must cause an *equal* flow of both surfaces. In the same way local high temperature must cause an *equal* melting of both surfaces. This means that both contribute to the welded joint; when the break occurs *both* surfaces will be distorted and torn.

An interesting confirmation of the view that welding takes place between sliding metals was obtained from some experiments with platinum sliding on silver. At the beginning of the run the behaviour was, as expected, that characteristic of dissimilar metals. Small stick-slips were observed, the friction was low, and a smooth groove was cut in the silver. After the slider had moved a short distance the behaviour changed to that characteristic of similar metals. The stick-slips disappeared, the friction rose to a high value, and the torn track now resembled that found in such cases. The results were similar to those reproduced in figs. 8*a*, 9*a*, and 10*a*. A close examination showed that a small blob of silver had become welded on to the upper surface so that the sliding was in reality silver on silver. Similar effects were observed with copper on silver. Again the sliding changed over to that characteristic of similar metals, and close examination showed that again silver had welded on to the copper surface.

The experiments show that three distinct types of metallic interaction and sliding may occur, but in practice the distinction between these types may not always be clear-cut. Under some conditions all these types of junctions may occur simultaneously. These observations have an obvious bearing on engineering practice, and the wear and friction of metals sliding under actual working conditions. It is clear that the friction and the damage to the surfaces will be less if the physical properties, particularly the melting-points of the two metals, are different so that junctions of *type 3* cannot occur. The most important surface, e.g. the journal, in a bearing should have a high melting-point and should slide on a surface of lower melting-point. The local high temperatures and high pressures will cause a welding together at the points of contact, but it is only the softer, lower melting metal which will flow. The surface of the higher melting metal will remain undamaged except that blobs of the other metal may adhere to it. These are the conditions which hold for a steel journal sliding on a bearing surface of white metal. In this case the surface temperature and the local pressure cannot rise above the values characteristic of white metal, so that the steel surface will not be affected.

If identical metals are slid together we may expect excessive wear and higher friction because the metals contribute equally to the welded junctions and both surfaces will be torn. Engineering experience shows that heavy wear does indeed occur if the two surfaces are of the same metal. If similar surfaces must be used they should be made of some non-homogeneous alloy so that the junctions formed are not those of *type 3*.

The fact that these different types of sliding occur is also important in the testing of lubricants. A lubricant which reduces one type of friction and wear may not necessarily be effective in reducing another. It is important that the surfaces of the test machine should be so chosen that the friction falls into one class or another.

We wish to express our thanks to the Department of Scientific and Industrial Research and to the Governing Body of Emmanuel College for a maintenance grant to one of us (L. L.), and to the Department of Scientific and Industrial Research and the Chemical Society for grants for apparatus. We are also indebted to Dr W. H. Hatfield for the gift of the special steel and to Professor R. S. Hutton for the use of apparatus in the Metallurgical Laboratory.

SUMMARY

A detailed analysis of the kinetic friction between moving metals shows that the frictional force does not remain constant during sliding. Sliding is not a continuous process: the motion proceeds in jerks and large and violent fluctuations occur in the friction. In the case of dissimilar metals the top surface sticks to the bottom one and moves with it until, as a result of the gradually increasing pull, a sudden and very rapid slip occurs. The process is then repeated indefinitely. A simultaneous measurement of the surface temperature shows that this too is fluctuating and, at the instant of slip, there is a sudden temperature "flash".

The exact behaviour depends upon the relative physical properties of the metals, particularly on the melting-point, and there is evidence that three distinct types of sliding may occur. The experiments suggest that friction is due to a welding together of the metals at the local points of contact. These metallic junctions are large compared with the dimensions of a molecule and when they are broken the metal is distorted to a considerable depth.

Even if the surfaces are lubricated with mineral oils or other lubricants an intermittent clutching and breaking away of the surfaces occurs and the behaviour may be essentially the same. Certain long chain fatty acids may prevent stick-slip and allow continuous sliding to take place.

REFERENCES

Amontons 1699 "Histoire de l'Académie Royale des Sciences avec les Mémoires de Mathématique et de Physique", p. 206.

Beare and Bowden 1935 *Philos. Trans.* A, **234**, 329.

Bowden and Leben 1937 "General Discussion on Lubrication and Lubricants. Institution of Mechanical Engineers. Group IV. Properties and Testing", p. 40.

— — 1938 *Nature, Lond.*, **141**, 691.

Bowden and Ridler 1935 *Proc. Camb. Phil. Soc.* **31**, 431.
— — 1936 *Proc. Roy. Soc.* A, **154**, 640.
Bridgman 1918 *Proc. Amer. Acad. Arts Sci.* **53**, 269.
— 1935 *Phys. Rev.* **48**, 825.
— 1936 *J. Geol.* **44**, 653.
— 1937 *Proc. Amer. Acad. Arts Sci.* **71**, 387.
— 1937 *J. Appl. Physics*, **8**, 328.
Coulomb 1785 "Mémoires de Mathématique et de Physique de l'Académie Royale des Sciences", p. 161.
De la Hire 1732 "Histoire de l'Académie Royale des Sciences", p. 104.
Derjaguin 1934 *Z. Phys.* **88**, 661.
Elam 1935 "Distortion of Metal Crystals", p. 147. Oxford Univ. Press.
— 1938 *Proc. Roy. Soc.* A, **165**, 568.
Euler 1750 "Histoire de l'Académie Royale des Sciences et Belles-Lettres", Annee 1748, Berlin, 1750, p. 122.
Ewing, Sir A. 1892 *Not. Proc. Roy. Instn.* **13**, 387.
— 1912 *J. Inst. Met.* **8**, 4.
Hardy, Sir W. B. 1936 "Collected Works". Camb. Univ. Press.
Hirn 1854 *Bull. Soc. Indust. Mulhouse*, **26**, 188.
Morin 1835 "Nouvelles expériences sur le frottement...faites à Metz en 1833." Paris.
Nadai 1931 "Plasticity." New York: McGraw-Hill Book Company.
Rennie 1829 *Philos. Trans.* **34**, 143.
Taylor and Farren 1925 *Proc. Roy. Soc.* A, **107**, 448.
Tomlinson 1929 *Phil. Mag.* **7**, 905.

The area of contact between stationary and between moving surfaces

By F. P. Bowden and D. Tabor
Laboratory of Physical Chemistry, Cambridge

(*Communicated by G. I. Taylor, F.R.S.—Received* 8 *August* 1938—*Revised* 29 *October* 1938)

[Plates 25, 26]

When two plane surfaces are placed together the area of intimate contact must be very much less than the apparent area. Even if the surfaces are very carefully polished and are made as flat as possible, hills and valleys will still be present on the surface. The upper surface will be supported on these irregularities, and large areas will be separated by a distance which is great compared with the dimensions of a molecule. We do not know very much about the size of these irregularities nor the degree of flatness of the surfaces. Optical methods cannot reveal irregularities much smaller than half a wave-length of light. The scattering of protons (Knauer and Stern 1929) or the diffraction of electrons (Thomson 1934; Finch, Quarrell and Wilman 1935) can give information about the structure of a small portion of the surface, but these methods would not reveal the presence of isolated peaks nor show how flat are the surfaces over a large area. Since it is difficult, even with the most refined technique, to prepare surfaces which are flat to within one or two thousand angstroms, we may expect that the area of intimate contact, that is, the area over which the surfaces are within the range of the molecular attraction, will, for most surfaces, be very small. Some knowledge of the real area of contact is essential for any complete understanding of the mechanism of friction, and this paper describes an attempt to estimate this area for both stationary and for moving surfaces.

1. THE AREA OF CONTACT BETWEEN STATIONARY SURFACES

In 1881 Hertz published his celebrated paper on the elastic deformation of solid surfaces and calculated how the area of contact between curved surfaces should depend upon the load. Bidwell (1883) measured the electrical conductance between crossed cylinders of carbon and bismuth, and Meyer (1898) made some careful measurements of the conductance of steel spheres in contact. Meyer assumed Hertz's equations to hold and deduced that the specific resistance of the steel within the area of contact was proportional to the pressure at that point. From the work of later investigators it appears that his surfaces could not have been clean. Further work on the resistance between polished plates was published by Aurén (1903) and Browning (1906). The next important contribution was made by Binder (1912), who showed that the conductance was smaller than would be expected if the whole area of the surfaces were touching. He suggested that real contact took place only over a small portion of the surface. Experimental verification of these observations was obtained by Pedersen (1916), although the interpretation he gave was different and was based on the assumption of a thin transitional layer of relatively high resistance. Support for Binder's views comes from the work of Holm and his collaborators who have published an exhaustive series of papers from 1922 onwards. Holm shows that the contact resistance between clean metals obeys Ohm's Law and is a "spreading resistance", i.e. it is produced as a result of the constriction of the current stream through a small area of contact. He concludes that for flat surfaces the contact occurs over a large number of small areas.

§ 1 of this paper describes the measurement of the area of contact between stationary surfaces by electrical and by visual means. Surfaces of different size, shape and material are used and the measurements are made over a wide range of loads, so that the results can be correlated with the frictional measurements described in the preceding paper.

General

Hertz has shown that, if elastic deformation is assumed, the theoretical radius of contact a between a spherical surface resting on a plane of the same material is given by

$$a = \sqrt[3]{\frac{3Wnr}{8E}},$$

where W = load between surfaces.

E = Young's modulus.

r = radius of curvature of spherical surface.

n = a numerical factor lying between 3 and 4 according as Poisson's ratio lies between $\frac{1}{2}$ and 0.

Hence a lies between the values

$$1{\cdot}08(Wr/E)^{\frac{1}{3}} \quad \text{and} \quad 1{\cdot}15(Wr/E)^{\frac{1}{3}}.$$

In what follows we shall use a value of

$$a = 1{\cdot}1(Wr/E)^{\frac{1}{3}}. \tag{1}$$

If two cylinders of the same material and the same radius r are pressed together with their axes at right angles, they will make contact over a circle the radius of which is also given by equation (1).

If now we consider a circular constriction of radius a in an infinite conductor of specific conductivity λ, the current is concentrated into the "neck" and spreads out on either side of it. This constriction of the current stream introduces a "spreading resistance" the value of which depends only on a and λ. A solution to a similar problem is given by Maxwell (1873) for the case of a cylindrical conductor of radius a with its extremity in metallic contact with a massive electrode. The spreading resistance involved is shown to be very nearly equal to $\frac{1}{4}a\lambda$. The spreading resistance R of the case considered above will be twice this, i.e.

$$R = \tfrac{1}{2}a\lambda. \tag{2}$$

This expression is, however, only strictly true for a current flowing through a plane circular constriction into a concentric sphere of infinite radius. If the conductor does not extend to infinity, the resistance will be less. If the conductor is a concentric sphere of radius r, the resistance will be approximately

$$R = \frac{1}{2\lambda}\left(\frac{1}{a} - \frac{2}{\pi r}\right), \tag{3}$$

when r is small compared with a; and more nearly

$$R = \frac{1}{2\lambda}\left(\frac{1}{a} - \frac{1}{r}\right), \tag{4}$$

when r tends to a. Thus so long as a is considerably smaller than r, the correction involved is small, and as subsequent measurements show is negligible.

Hence if two conductors make metallic contact over a circle of radius a, where a is small compared with the dimensions of the conductors, the electrical conductance Λ of the "junction" can be written with sufficient accuracy as

$$\Lambda = 2a\lambda. \tag{5}$$

From equations (1) and (5) $\Lambda = k_1 W^{\frac{1}{3}}$, (6)

where k_1 is a constant, so that if we assume *elastic deformation* the electrical conductance should be proportional to the *cube root* of the load.

If we assume *plastic deformation* we may consider that the surface irregularities will flow under the applied load until the cross-section of the contact is sufficient to support the load. In this case

$$W = \pi a^2 f. \tag{7}$$

Where f is the maximum pressure the metal can withstand without flowing. Combining equations (5) and (7) it follows that

$$\Lambda = k_2 W^{\frac{1}{2}}, \tag{8}$$

where k_2 is a constant, i.e. the electrical conductance should be proportional to the *square root* of the load.

If the metal is very soft and the pressure is great, the diameter of the contact area may be of the same order of magnitude as the dimensions of the metal surfaces forming the contact. In this case the spreading resistance is more nearly represented by equation (4) where it is seen that when $a = r$ the spreading resistance is zero. We can rewrite equation (4) as

$$\Lambda = 2\lambda \frac{ar}{r-a}. \tag{9}$$

Experimental

The first experiments were carried out on the contact formed between two crossed cylinders with their axes at right angles. The electrical conductance Λ of the junction was measured by a current-potential method. A known current i was passed through the contact, and the potential v was measured with a potentiometer or sensitive galvanometer of relatively high resistance, the arrangement being as shown in the accompanying diagram (fig. 1). Since separate current and potential leads were used, the conductance is given by $\Lambda = i/v$. The lines of current flow are also indicated. These are not the same as for a constriction at the centre of a spherical conductor, but so long as the constriction is small compared with the radius of the cylinders, the difference does not appear to be very great and is neglected.

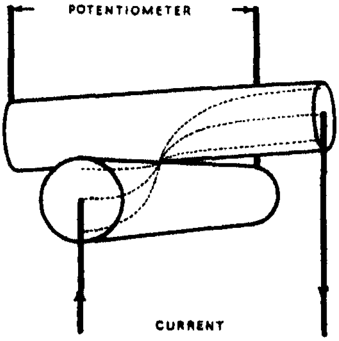


Fig. 1. Diagram showing arrangement of cylinders.

Subsidiary experiments showed that the conductance thus measured was independent of the current flowing, provided that the potential across the contact did not exceed a certain critical value. This aspect of the problem has been dealt with exhaustively by Holm (1929).

The load on the surfaces was varied by running water into or syphoning it out of a vessel to which the cylinders were suitably attached. The load range was 20–600 g. Using mercury instead of water the range was extended to about 6000 g. For large loads a special lever machine was used. The load was applied by means of a spring balance and could be increased to 1000 kg. The surfaces were cleaned and prepared in various ways, by fine grinding, by polishing, by etching with various reagents and by scraping with a sharp edge. The method of preparation did not have much influence on the results except that the scraped surfaces seemed to give more reproducible results.

Results

When the metal surfaces were placed together with a light load the conductance was at first variable. If a vibrating tuning-fork was held against the surfaces the conductance increased slightly and reached a steady state which was comparatively reproducible. This technique was first developed by Meyer (1898). For heavy loads this procedure was not effective, but a slight relative movement of the surfaces produced the same effect. It is evident that this process causes a better fitting together of the two surfaces and probably crushes out contamination at the region of contact.

The effect of load and the nature of the metal on electrical conductance

If the clean surfaces are placed together with a heavy load and are subjected to vibration, the value of the electrical conductance becomes reasonably reproducible. If the load is gradually reduced the conductance decreases in a uniform manner. If now the load is again increased (without vibration) there is only a small increase in the conductance, and further cyclic variations in load produce successively less change in the conductance. The reason for this hysteresis is not clear, but it is possibly due to the successive formation of oxide on the surface which may occur during the tearing of the metallic junctions (see later). If the load is again applied and the surfaces subjected to vibration, the conductance again rises to its former value. The following measurements were made on the reproducible part of the curve, i.e. the initial heavy load was applied to the freshly cleaned surfaces, and the surfaces were subjected to vibration. The load was then decreased gradually and the electrical conductance observed.

Experiments showed that the absolute value of the conductance for individual measurements in different experiments might vary in the case of heavy loads by 10–20 %. In the case of light loads the variation might be still greater. The slope of the curves was, however, much more reproducible. The results obtained for crossed cylinders of various metals and for carbon are shown in fig. 2. Each curve represents the mean of a large number of determinations. On the same figure the specific conductivities of the metals are plotted and it will be seen that:

(*a*) The variation of conductance with load for decreasing load follows a simple law of the type $\Lambda = kW^{1/n}$.

(*b*) The position of the curve for each material is in the order of the specific conductivity of the material.

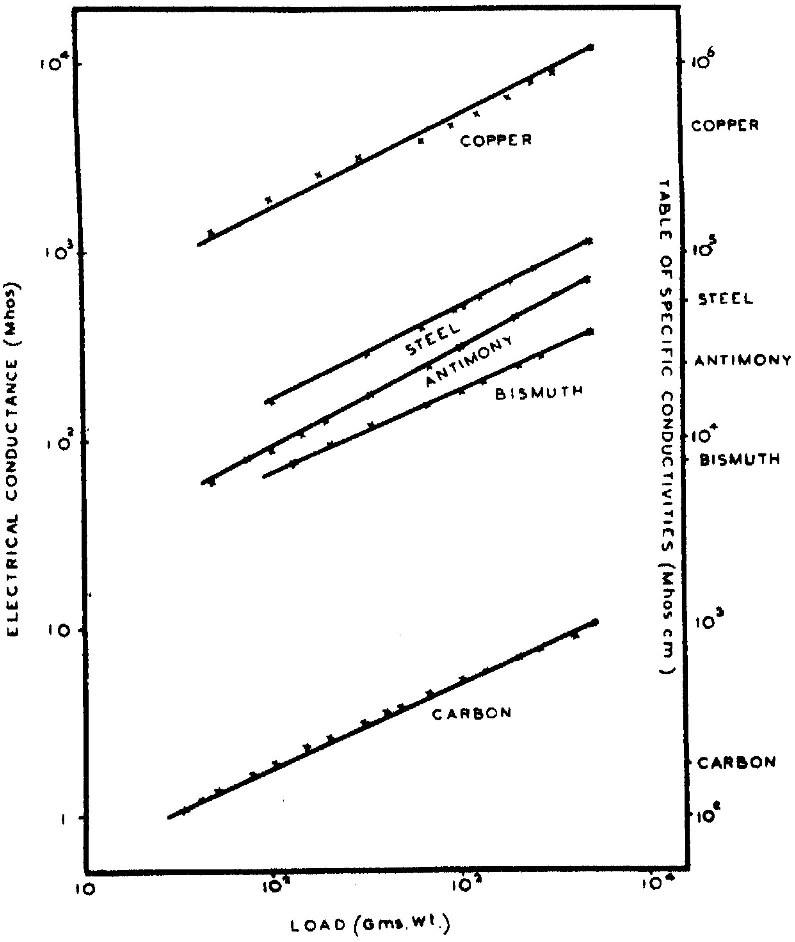


FIG. 2

Comparison of experimental results with theoretical curves for plastic and elastic deformation

In fig. 3, curve I, the experimental results for crossed cylinders of copper and steel are shown in more detail. The theoretical values of the conductance calculated on the assumption of elastic deformation (curve II) and plastic deformation (curve III) are also plotted on the same figure for comparison. In these cases the radius of contact is small compared with the diameter of the cylinders and equation (5) is used. It is seen that the general agreement between theoretical and experimental results is quite good, and that the observed slope of the curve corresponds more closely to *plastic* than to *elastic* deformation.

Some results for crossed cylinders of silver under heavy loads are given in fig. 4. In these experiments the radius of contact a was actually measured by visual means. After each load had been applied the cylinders were separated and the diameter of the impression determined with a microscope. Since at these heavy loads the radius of the contact was not negligible compared with the radius of the cylinders, equation (9) was used to calculate the conductance in terms of a. These values are plotted on curve II, whilst the experimental values of the conductance are plotted on curve I. It will be seen that the agreement between experiment and theory is again quite good.

The influence of the size and shape of the surfaces on the electrical conductance

According to the empirical theory of plastic deformation, a spherical surface resting on a flat one should have essentially the same conductance as two crossed cylinders, and experiments were carried out to test this. The results for steel are plotted in curve I, fig. 5. The crosses refer to crossed cylinders and the circles to a sphere on a flat. The spherical surface had an initial radius of curvature of 0·3 cm. During the experiment some flattening occurred, so that the radius of curvature increased to several cm., but this had no appreciable influence on the load-conductance curve. It will be seen that the results for both types of contact are essentially the same: the values lie on top of one another and the slope is again one half.

The results for flat steel surfaces are shown in the same figure (curves IIa, IIb). These surfaces were not polished: they were finely ground and were flat to a few fringes. They were then cleaned with rouge and water, caustic soda, water and acetone. Curve IIa is for a flat surface 0·8 sq. cm. in area, curve IIb for a flat surface 21 sq. cm. in area. The flat disks were placed in contact, a heavy load applied, and the disks were then given a sharp tap in a direction parallel to the surface. The conductance rose to a fairly reproducible value. The load was then gradually reduced and the conductance observed. The results were fairly reproducible and the points plotted are the mean of about twenty runs.

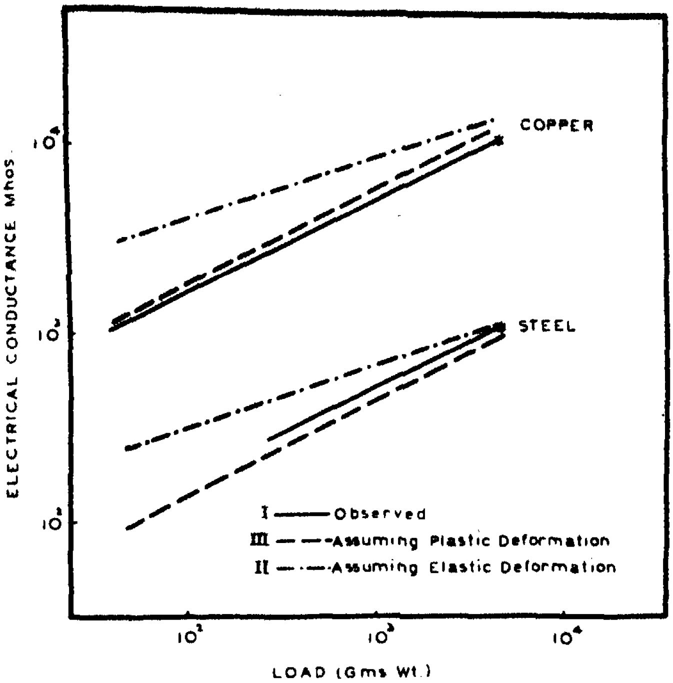


FIG. 3

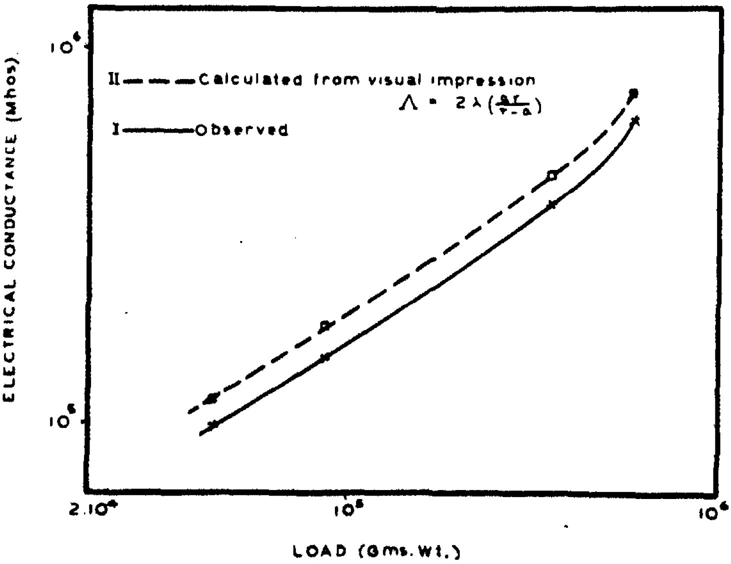


FIG. 4. Silver contacts. Crossed cylinders (diam. $2r = 0{\cdot}9$ cm.)

The results for flat surfaces do not lie on a straight line, but it is clear that the values of the conductance are not very different from those observed with curved surfaces.

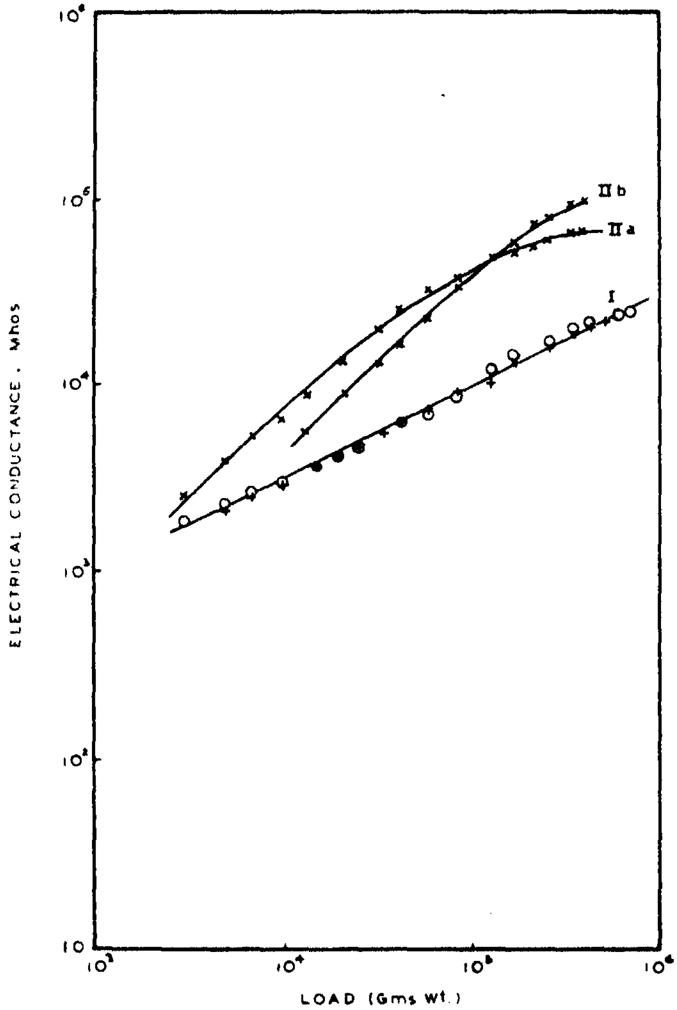


FIG. 5

Magnitude of the area of contact

We may form an estimate of the actual area of contact in several ways:

(i) From the observed value of the electrical conductance. The area of contact is then given by πa^2, where a is obtained from equation (9).

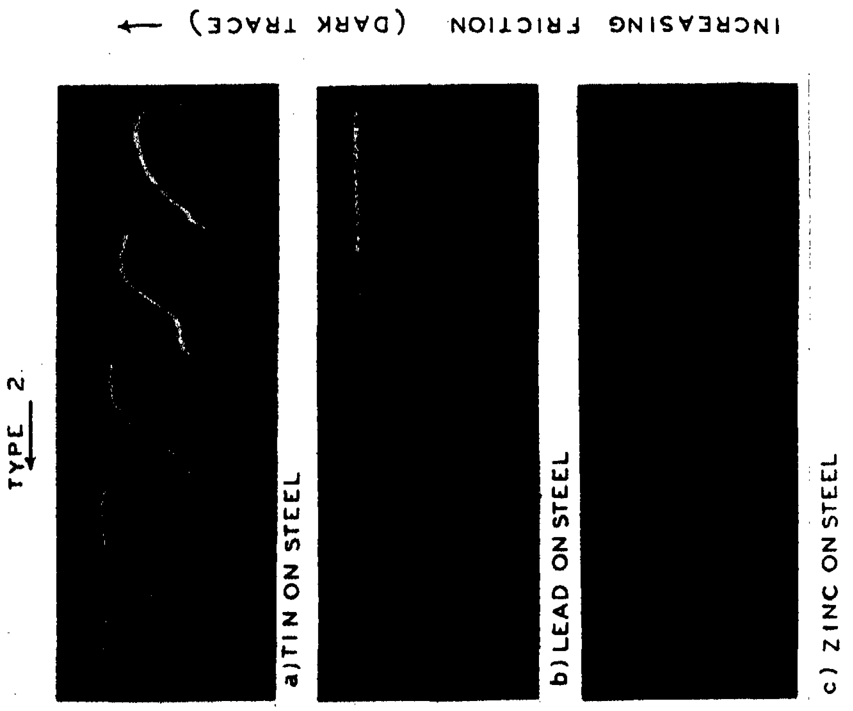

INCREASING FRICTION (DARK TRACE) →
TYPE 2
a) TIN ON STEEL
b) LEAD ON STEEL
c) ZINC ON STEEL

Fig. 7.

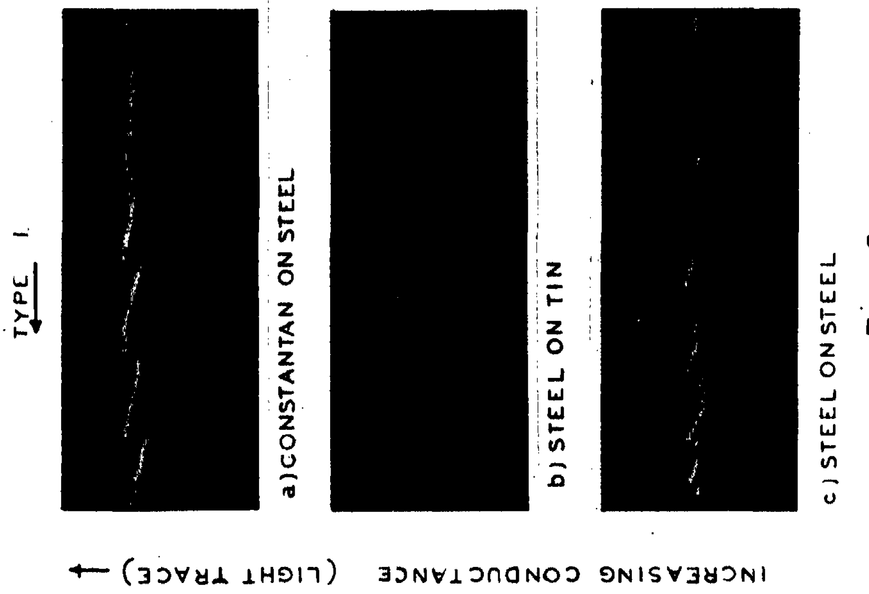
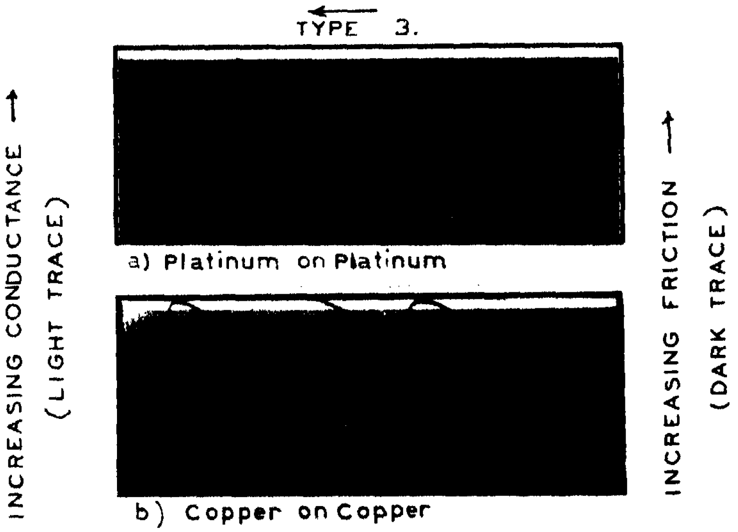

TYPE 1
a) CONSTANTAN ON STEEL
b) STEEL ON TIN
c) STEEL ON STEEL
INCREASING CONDUCTANCE (LIGHT TRACE) →

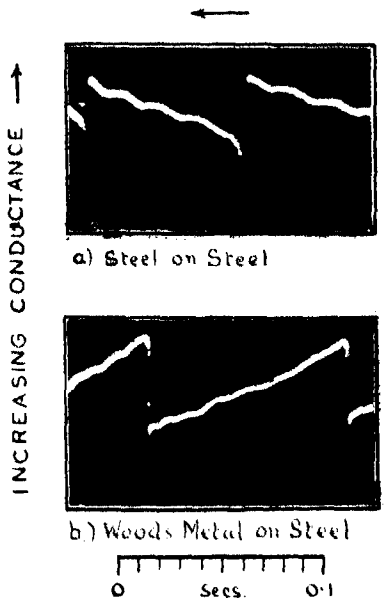
Fig. 6.



a) Platinum on Platinum

b) Copper on Copper

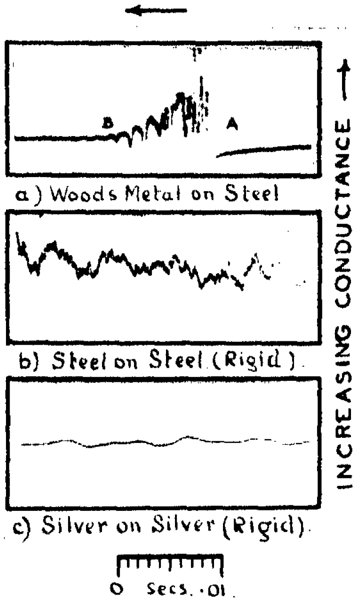
FIG. 8



a) Steel on Steel

b) Woods Metal on Steel

FIG. 9



a) Woods Metal on Steel

b) Steel on Steel (Rigid)

c) Silver on Silver (Rigid)

FIG. 10

For small loads equation (5) is sufficiently accurate, but above a load of about 50 kg. the correction becomes appreciable and equation (9) must be used.

(ii) From microscopic examination of the impressions left on the surfaces of the contacts.

(iii) From calculations assuming plastic flow. Here we can assume that equation (7) holds. This involves a knowledge of the flow pressure of the metal. We do not know this with certainty, but some data are available (Taylor and Farren 1925; Nadai 1931; O'Neill and Greenwood 1932).

Approximate calculations by these three independent methods are given in Table I for contacts of crossed cylinders, of steel and silver. The conductivity of a metal is of course dependent on the pressure and the relation is given by the empirical expression $\lambda_p = \lambda_0(1 + n \times 10^{-9}p)$, where the pressure p is expressed in g./cm.^{2}. For silver $n = 3{\cdot}6$ and for steel n is probably of the same order. Even at the region of contact where the pressure reaches the flow pressure of the material, the specific conductivity is altered only by a few per cent. The effect of this on the conductance of the junction is less than the experimental error and is neglected.

TABLE I

Load kg.	Observed conductance mhos	Experimental area of contact sq. cm.		Calculated area of contact sq. cm. plastic flow
		Electrical	Visual	
Steel contacts: $r = 0{\cdot}5$ cm.; $f = 10^7$ g./cm.^{2}; sp. cond. $\lambda = 75{,}000$ mhos cm.				
1	1,000	0·00013	—	0·0001
5	2,200	0·00061	—	0·0005
50	6,200	0·0045	0·0045	0·005
500	22,700	0·042	0·045	0·05
Silver contacts: $r = 0{\cdot}45$ cm.; $f = 2{\cdot}5 \times 10^6$ g./cm.^{2}; sp. cond. $\lambda = 620{,}000$ mhos cm.				
0·5	10,000	0·0002	—	0·0002
5	33,000	0·002	—	0·002
50	111,000	0·018	0·019	0·02
500	550,000	0·15	0·19	0·2

Influence of the physical state of the surface on the electrical conductance.

Some attempts were made to investigate the influence of the physical state of the surface on the electrical conductance and hence on the area of contact. The surface of the steel cylinders was rubbed with a coarse file, so that it was as rough as possible and the conductance was compared with that of a very finely ground surface. No appreciable difference between the two was observed.

Discussion

An examination of fig. 1 shows that the measured values of the conductance between metal surfaces lie in the same order as the specific conductivities of the metals. Since the mechanical properties are not widely different, this suggests that the measured resistance is essentially metallic (or carbon/carbon) and is not that of oxides or contaminating films on the surface. It would be an odd coincidence if the electrical conductivities of the oxides or other surface films stood in exactly the same order as the conductivity of the metals. It is clear that if contaminating films are present between the surfaces they must be so thin that their contribution to the electrical resistance is very small.

The results show quite clearly that the conductance varies in an orderly manner with the applied load. If we assume *elastic* deformation of the metal to occur at the contact we know from equation (6) that the conductance should be proportional to the *cube root* of the load. On the other hand, equation (8) shows that if *plastic* flow occurs the conductance should be proportional to the *square root* of the load. Owing to the uncertainties involved in the conductance measurements it is not always easy to distinguish between these two possible characteristics. Fig. 3 does, however, indicate that the variation of conductance with load is in much better agreement with *plastic* than with elastic deformation. This is shown most clearly for high loads on contacts of silver in fig. 4, but the fact that it is valid for small loads too, means that even for loads of the order of 40 or 50 g. the metal flows.

For surfaces which, by their geometric form, make contact in a single region (e.g. crossed cylinders, or sphere-on-flat) the measured values of the conductance suggest that intimate contact occurs over the whole of the deformed area. This is shown very clearly in Table I. The area of contact between crossed cylinders has been calculated assuming plastic flow. This value is compared with experimental values obtained in two quite different ways—from measurements of the electrical conductance and from the size of the visual impression. It will be seen that the area of contact deduced from these methods is in surprisingly good agreement with the calculated value. In fig. 5, curve I, contacts composed of crossed cylinders and sphere-on-flat are compared. Though for the latter contact the radius of curvature of the spherical surface varied by a factor of at least ten, the absolute values of the conductance and the slope of the curves are almost identical. We can conclude that, provided the actual contact is localized in a single continuous area, the conductance and hence the real area of contact will be reasonably independent of the shape and radius of curvature of the surfaces.

The results for flat surfaces are shown in curves II*a* and II*b*, fig. 5. It will be seen that, although the possible area of contact is enormously greater than that of the curved surface, the observed value of the conductance is of the same order of magnitude. At a load of 5 kg. for example, the area of contact of the curved steel contact is 5×10^{-4} sq. cm. If nearly all the surface of the 21 sq. cm. flat were in contact, the area of contact and hence the conductance, would be *ca.* 40,000 times as great. Experiment shows in fact that it is only twice as great. It is thus clear that only a small fraction of the area of the flat surfaces can be in intimate contact: they must touch at a number of legs or bridges. It will also be noted that the conductance of both sets of flats is almost the same, although their apparent areas stand in a ratio of 30 : 1. We can conclude that the conductance is almost independent of the apparent area of the surfaces: it depends mainly on the load.

It is difficult in the case of flat contacts to make an exact estimate of the real area of contact from the conductance measurements alone. The value of the conductance must depend both on the size of the metallic bridges and on their number. Since the spreading resistance of each bridge is inversely proportional to its diameter, and the area of contact is proportional to the square of the diameter, it follows that, for a given conductance the area of contact is inversely proportional to the number of bridges. We do not know this number with any certainty, though in the case of flat surfaces it is clear that the number of legs on which they rest cannot be less than three. Experiments suggest in fact, that the number of points of contact is greater than this. Thus when the surfaces were separated and examined, a very large number of small marks were observed distributed over the whole area. It is not certain that contact occurred simultaneously at all these points, but as a conservative estimate we may assume that the surfaces were resting on ten legs. On this basis we can calculate the fraction of the apparent area really in contact, at any given load. The following results are for the 21 sq. cm. flats.

TABLE II

Load kg.	Fraction of macroscopic area in contact
300	1/130
100	1/700
20	1/10,000
3	1/170,000

It is thus clear that for flat surfaces the real area of contact even for appreciable loads is a small fraction of the apparent area. The flat surfaces are held apart by small surface irregularities which form bridges of an essentially metallic nature. These bridges flow under the applied pressure or their number increases until their total cross-section is sufficient to enable them to support the applied load.

In general the load conductance curve for flat contacts appears to be steeper than for curved surfaces, but it is difficult to put forward a satisfactory quantitative hypothesis to explain this.*

2. THE AREA OF CONTACT BETWEEN MOVING SURFACES

The measurements were extended to moving surfaces. The method of measuring the conductance was similar, except that the potentiometer was replaced by an instrument of rapid period such as an Einthoven galvanometer or a cathode-ray oscillograph. The friction was measured by a high-frequency apparatus similar to that described in the preceding paper. The lower surface was in the form of a flat plate and was driven at a uniform rate by a water piston. The upper surface which rested on it was usually curved. The conductance and the friction measurements were both photographed by the same moving-film camera, so that a simultaneous record of both quantities was obtained. The lower surfaces were ground flat and cleaned by the methods previously described.

Results

(1) *Higher melting metal on lower melting metal.*

Fig. 6*a* (Plate 25) shows the results obtained for a curved surface of constantan sliding on a steel plate. The dark trace represents friction and the white trace the electrical conductance. The direction of motion is from right to left as shown by the arrow. The load was 430 g., and the speed 0·005 cm./sec.

Fig. 6*b* (Plate 25) was obtained with an upper surface of steel sliding on a flat tin surface. Other combinations, molybdenum on steel, nickel on steel, platinum on steel, steel on tin, lead on tin, platinum on silver, were used and the general result was the same, although the magnitude of the slip and of the conductance change varied with the nature of the metals. Fig. 6*c* (Plate 25) shows the results for steel on steel.

---

* See, for example, Holm's ingenious hypothesis of a "sieve-resistance". This would explain a slope of 1 for flat contacts. The absolute values would, however, be considerably different from the values he observes.

(2) *Lower melting metals on higher melting metals.*

In this case the upper curved contact had a melting point less than that of the lower flat surface. The load and the speed were of the same order. Fig. 7*a* is for tin on steel, fig. 7*b* for lead on steel, and fig. 7*c* for zinc on steel (Plate 25). Other combinations of Wood's metal on steel and Wood's metal on nickel were used and gave similar results.

(3) *Similar metals.*

Both the upper and the lower surfaces were made of the same metal. The results for platinum on platinum are shown in fig. 8*a* and for copper on copper in fig. 8*b* (Plate 26). Combinations of tin on tin, silver on silver and nickel on nickel gave similar results.

Discussion

The change in area of contact during sliding.

It is clear from these experiments that the area of contact between moving surfaces does not remain constant. The average value of the conductance is not very different from that observed at stationary surfaces, but large fluctuations may occur in it during sliding. These changes show a close correlation with the frictional stick-slip, and the exact behaviour depends upon the nature of the metals. In the case of a curved slider of high-melting metal sliding on a low-melting metal (figs. 5*a*, *b*) the area of intimate contact *increases* during the stick. The moment slip occurs there is a sudden *decrease* in this area. The results lend strong support to the suggestion put forward in the earlier paper, that the friction under these conditions (*type* 1) is due primarily to the small irregularities on the high-melting metal ploughing through the surface of the softer metal. The characteristic cut scratch is observed with this type of behaviour. Steel on steel behaves in a similar manner though this anomaly is probably due to its inhomogeneous nature (see preceding paper).

In the case of a low-melting metal sliding on one of high melting-point the behaviour is the opposite of this. An examination of figs. 7*a*, *b*, *c* shows that the conductance *decreases* during the stick. The moment after slip occurs it *rises* to a maximum value. This supports the suggestion that in this case (*type* 2) the local high pressure aided by the temperature rise caused by the frictional heat, welds or solders the lower-melting metal on to the surface of the higher melting one: the area of contact is then at a maximum. During the stick, the increasing pull will draw out these soldered junctions thinner and thinner (conductance decreasing) until they suddenly break and slip occurs. The process is then repeated.

When both surfaces are of the same metal, the conductance measurements show that the area of contact is somewhat greater and it remains much more constant during sliding. An examination of fig. 8*a* (Plate 26), for example, shows that the average conductance between two platinum surfaces is about 1200 mhos as compared with about 100 mhos for platinum on steel. Only small variations occur in this value during sliding. The friction is high and shows slow fluctuations and the characteristic torn track is formed. This behaviour is generally observed with similar metals provided they are homogeneous. Again, these conductance measurements lend support to the suggestion that in the case of similar metals (*type* 3) a mutual welding of the two surfaces takes place. The local high pressures and temperatures will cause a flow of both surfaces which now contribute *equally* to the welded junctions. These mutual junctions will probably be more easily formed, and the real area of contact and the friction will tend to be greater. When movement occurs both surfaces will be damaged and the metal in the resulting track will be badly distorted and torn.

Calculation of area of contact.

If we consider, in the case of a curved surface sliding on a flat plate, that the maximum conductance Λ_m corresponds to a single circular metallic junction of diameter $2a$, then

$$\Lambda_m = 4a\left(\frac{\lambda_1\lambda_2}{\lambda_1+\lambda_2}\right)$$

or

$$2a = \frac{\Lambda(\lambda_1+\lambda_2)}{2\lambda_1\lambda_2},$$

where λ_1 and λ_2 are the specific conductivities of the two metals. We may also form an estimate of $2a$ by direct microscopic measurement of the width of the track produced on sliding.

In calculating $2a$ from the conductance, the normal values of the electrical conductivity were used. As mentioned earlier, the pressure effect on λ is very small and may be neglected. The effect of temperature, however, is considerably greater. Thus with most metals a rise in temperature of 10° C will decrease the conductivity by a few per cent. It is very difficult to make an accurate correction for this effect because the temperature distribution at the surface irregularities is not known with certainty. Experiments have shown however that the high temperature is confined to a very thin layer on the surface (Bowden and Ridler 1936) so that the effect on the measured conductance is probably small. The calculated values of $2a$ can only be regarded as approximations. The temperature effect would if anything increase these values, and this would tend to bring them into better agreement with the visual observations.

Table III shows some values of $2a$ calculated from the conductance and from visual measurements. The table is divided into three groups which correspond to the three types of sliding. The load was roughly the same in each case (about 400 g.) and the sliding speed *ca* 0·01 cm./sec. It will be seen that when the upper curved surface is of the higher-melting metal (*type* 1) the agreement between the electrical and the visual measurements of the diameter of contact is quite good. The values are given in columns 3 and 4.

TABLE III

	A_m mhos	$2a$ cm. Calc. from A_m	$2a$ cm. Visual track width	Area of contact A sq. cm. Calc. from A_m
		Type 1		
Steel on tin	60	10^{-3}	$2\text{–}3 \times 10^{-3}$	$0{\cdot}01 \times 10^{-4}$
Nickel on steel	200	3×10^{-3}	4×10^{-3}	$0{\cdot}07 \times 10^{-4}$
Platinum on steel	150	2×10^{-3}	$1\text{–}2 \times 10^{-3}$	$0{\cdot}03 \times 10^{-4}$
Molybdenum on steel	330	4×10^{-3}	10×10^{-3}	$0{\cdot}12 \times 10^{-4}$
Steel on copper	700	8×10^{-3}	10×10^{-3}	$0{\cdot}5 \times 10^{-4}$
Steel on steel	100	$1{\cdot}7 \times 10^{-3}$	$2\text{–}3 \times 10^{-3}$	$0{\cdot}03 \times 10^{-4}$
		Type 2		
Tin on steel	200	3×10^{-3}	50×10^{-3}	$0{\cdot}07 \times 10^{-4}$
Zinc on steel	100	$1{\cdot}1 \times 10^{-3}$	60×10^{-3}	$0{\cdot}01 \times 10^{-4}$
		Type 3		
Platinum on platinum	1200	13×10^{-3}	18×10^{-3}	$1{\cdot}2 \times 10^{-4}$
Copper on copper	5000	9×10^{-3}	12×10^{-3}	$0{\cdot}6 \times 10^{-4}$
Nickel on nickel	1000	7×10^{-3}	10×10^{-3}	$0{\cdot}4 \times 10^{-4}$
Silver on silver	6000	10×10^{-3}	17×10^{-3}	$0{\cdot}8 \times 10^{-4}$

When the upper curved surface is of a lower-melting metal, however, the agreement is very bad. The observed track width is considerably wider than that calculated from the conductance. The large disparity between these two quantities is characteristic of this type of contact and is to be expected. The conductance depends upon the real area of contact, whereas the track width merely corresponds to a broad smear within which the real points of contact occur.

In the case of similar metals (*type* 3) A_m is somewhat greater than that previously observed, and the values of $2a$ calculated by both methods are again in good agreement. The last column of the table shows the real area of contact calculated from the conductance measurements. These values are necessarily only approximate as already pointed out. It will be seen that at this load the real area of contact lies between one ten-thousandth and one-millionth of a square centimetre. It tends to be greater for surfaces of *type* 3 than for surfaces of *types* 1 and 2.

The sliding of flat surfaces.

In order to determine how far the behaviour was influenced by the shape of the surface, experiments were carried out with flat surfaces. The upper curved surface was replaced by a flat disk *ca.* 1 cm. in diameter. Experiments were made with tin on steel and steel on steel. The results were essentially the same as for a curved contact. The friction, the electrical conductance and the variations in these quantities were almost identical with those observed with a curved slider. Further experiments were carried out with large flat surfaces of steel 21 sq. cm. in area. Although the possible area in this case was some million times as great as that for a curved surface the actual value of the conductance was almost the same.

An interesting confirmation of this observation is given by measurements of the surface temperature of sliding metals. Bowden and Ridler (1936) have shown that when metals slide on one another the surface temperature may reach a high value. Some unpublished experiments (Ridler 1936) showed that the actual temperature reached is almost independent of the apparent area of the surfaces. Thus when the load and speed were kept constant (80 g. and 150 cm./sec. respectively) the recorded temperature of metal surfaces 0·008 sq. cm. in apparent area was 120° C. When the area was increased to 0·64 sq. cm. it was 90° C. It is seen that although the possible area of contact had been increased by a factor of 80, the surface temperature rise was nearly as great. This suggests that the area at which the heat is liberated, that is the area of intimate contact, is approximately the same in both cases.

Analysis of the slip.

The measurements so far described were carried out with an Einthoven galvanometer, so that the observations were limited by the period of the instrument, which was approximately one-thousandth of a second. When a cathode-ray oscillograph was used the measurements could be extended to events happening in less than one hundred-thousandth of a second, so that a more detailed analysis of the conductance changes was possible.

A typical photograph showing in more detail the behaviour of metals of *type* 1 during slip is shown in fig. 9*a* (Plate 26). The time base was relatively slow so that the record is not a very high-speed one. On the photograph 0·1 mm. corresponds to 0·00036 sec. The ripples are instrumental and should be ignored. It is clear that when slip occurs the conductance change is much greater than that previously recorded. The conductance which before slip is 100 mhos falls to a low value of 50 mhos and then rapidly rises to the value of 80 mhos recorded by the Einthoven.

The behaviour of surfaces of *type* 2 is also interesting. Fig. 9*b* (Plate 26) is for Wood's metal on steel. The previous record showed that, after slip, the conductance rose to a maximum value. This more detailed record shows that when slip occurs the conductance first *falls* to a low value and then rises rapidly. This is what we should expect on the mechanism suggested. When slip occurs the welded junctions are broken, so that the area of metallic contact, and hence the conductance first *falls*, and then rapidly rises when the new junctions are formed. A still more detailed analysis with a rotating drum camera running at high speeds is shown in fig. 10*a* (Plate 26). The distance AB now corresponds to the time of a single slip, in the case of Wood's metal on steel. It is clear that the behaviour is very complex and that subsidiary and very rapid clutching and breaking away of the surfaces is occurring during the slip. A close examination of the photograph shows that fluctuations in the conductance are occurring in a time interval less than 10^{-5} sec. When similar metals were used only small fluctuations were observed.

Experiments with rigidly-held surfaces.

In the experiments so far described the upper surface was attached to a high-frequency measuring device, and the restraint was such that a momentary arrest could occur. Under many practical conditions, for example in most parts of an engine, any arrest is difficult because one surface is rigidly held and the other is driven steadily on. It is important to determine whether or not under these conditions the same intermittent clutching and breaking away of the surfaces takes place. For this purpose the friction-measuring apparatus was removed and the upper surface was fixed to one end of a short steel bar of ½ in. square cross-section. The other end of the bar was attached to a ball bearing, so that the load could be applied by adding weights to the bar. Thus the upper surface could move freely in a vertical plane but was held rigidly in the direction of motion. The friction was not measured but the electrical conductance was recorded on a cathode-ray oscillograph. Experiments were carried out with Wood's metal on steel, steel on steel and silver on silver. The load could be varied from 50 to 2000 g. A typical result for steel on steel with a load of 60 g.

and a surface speed of 0·025 cm./sec. is shown in fig. 10*b* (Plate 26). It is clear that variations in the electrical conductance and hence in the area of contact still occur during sliding. These variations are smaller than those observed with the first type of restraint and they are very rapid. On the photograph 0·1 mm. corresponds to 7×10^{-5} sec. The sliding speed is 0·025 cm./sec., so that the relative movement of the surfaces during this time is only 10^{-6} cm. Although this relative movement corresponds only to the distance of a few atoms the measurements show that a clutching and breaking away occurs. We should expect that the length of the stick would be controlled by the elasticity of the moving system and in the limiting case by the elasticity of the surface irregularities themselves.

Amontons's law.

There has been considerable difficulty in explaining the experimental observations, first made by Amontons, that the tangential frictional force is directly proportional to the normal load. It was usual to assume that elastic deformation of the solids occurred, so that the real area of contact A should vary as the two-thirds power of the normal load. A good discussion of this has been given recently by Adam (1938), who points out that on such an assumption Amontons's law would imply that the tangential frictional force must increase in proportion to $A^{\frac{3}{2}}$, "a result so unlikely that the meaning of the law is still mysterious".

It is clear from these measurements, however, that the deformation of the surface irregularities is not elastic: it is plastic. The real area of contact is directly proportional to the normal load. The difficulty then disappears. The total cross-section of the junctions and hence the tangential force required to break them will be directly proportional to the applied load, and we should expect Amontons's law to hold true.

In the same way these measurements offer an explanation of Amontons's other observation, that the friction is independent of the macroscopic area of the surfaces. Since the area of intimate contact is nearly independent of the apparent area we should naturally expect that the friction would also be independent of the size of the surface.

Lubricated surfaces.

Some preliminary experiments were made with steel surfaces lubricated under boundary conditions. With mineral oils and pure hydrocarbons the electrical conductance was of the same order of magnitude as that observed with unlubricated surfaces, indicating that the conductance was essentially metallic. When the surfaces were set in motion, the same characteristic fluctuations in the friction and conductance occurred as for unlubricated metals. The fall in conductance during the slip was, however, considerably greater: in some cases it fell almost to zero. This suggests that during the slip a relatively thick wedge of oil may form between the surfaces. When long-chain fatty acids were used, so that continuous sliding occurred, the conductance was still high, but it showed little variation.

The pressure on boundary films.

The fact that the real area of contact is so small also has important practical implications. Even with lightly loaded surfaces the local pressure between unlubricated surfaces may be sufficiently great to cause the flow of steel. When large surfaces are used it does not necessarily mean that the real pressure is much less but merely that the points of contact are more widely distributed.

If the surfaces are separated by a thick layer of oil so that the conditions are those of fluid lubrication the pressure will, of course, be inversely proportional to the area of the surfaces. Under boundary lubrication conditions, however, this will not be the case since the load is carried by the surface irregularities. In this sense all boundary lubricants may be extreme pressure lubricants. Unless the lubricant molecule possesses an active group capable of attaching it very firmly to the surface, the local high pressures will force it from underneath the points of contact. When very heavy loads are used (the so-called "extreme-pressure" conditions) the local pressure may not in reality be very much greater since it cannot exceed the flow pressure of the metal. The frictional heat liberated, however, and the local surface temperatures may under these conditions be very great. As shown elsewhere (Bowden and Ridler 1936) "extreme-pressure" lubricants should perhaps be called "extreme-temperature" lubricants.

The work described above was carried out as part of the programme of the Lubrication Research Committee and is published by permission of the Department of Scientific and Industrial Research. The authors wish to express their thanks to the Royal Society and to the Chemical Society for grants for apparatus.

SUMMARY

An estimate of the real area of contact between stationary and between moving metal surfaces has been made by electrical and by visual means. Experiments with stationary surfaces show that the area of intimate contact is very small. It varies with the pressure, but for flat steel surfaces

it may be less than one ten-thousandth of the apparent area. The real area of contact is not greatly affected by the size, shape and degree of roughness of the surface: it depends mainly on the pressure. The general behaviour is consistent with the view that the surfaces are held apart by small irregularities. This means that even with lightly loaded surfaces the local pressure may be sufficiently great to cause steel to flow plastically. This conclusion is supported by the observed relation between the pressure and the electrical conductance.[1] The slope of such a curve indicates that, at the points of contact, plastic flow of the metal has occurred. Although the stresses may cause elastic deformation in the bulk of the metal, the experiments suggest that the irregularities on which the bodies are supported are crushed down until their cross-section is sufficient to enable them to support the applied load.

Measurements made with moving surfaces show that the area of contact is not constant but is fluctuating rapidly during sliding. A detailed analysis shows a remarkable correlation between these fluctuations and the frictional and temperature changes described in an earlier paper. It is clear that an intermittent clutching and breaking away of the surfaces is taking place and the results support the view that metallic junctions between the metals are being rapidly formed and broken. The nature of these junctions depends on the relative physical properties of the two metals and there is again evidence that three main types may occur.

The conductance measurements show that even if the metals are lubricated with mineral oils and other lubricants, metallic contact may still occur through the film of lubricant. When the surfaces slide, fluctuations in the area of contact are observed and the behaviour may be essentially the same as for unlubricated surfaces.

REFERENCES

Adam, N. K. 1938 "The Physics and Chemistry of Surfaces", p. 232. Oxford Univ. Press.

Aurén, T. E. 1903 *Ark. Mat. Astr. Fys.* **1**, 25–41.

Bidwell, S. 1883 *Proc. Roy. Soc.* A, **35**, 1–19.

Binder, L. 1912 *Elektrotech. u. Maschinenb.* **38**, 781–8.

Bowden, F. P. and Ridler, K. E. W. 1936 *Proc. Roy. Soc.* A, **154**, 640–56.

Browning, W. 1906 *J. Instn Elect. Engrs*, **37**, 372–4.

Finch, G. I., Quarrell, A. G. and Wilman, H. 1935 *Trans. Faraday Soc.* **31**, 1050–80.

Hertz, H. 1881 *J. reine angew. Math.* **92**, 156–71. Also "Collected Works."

Holm, R. 1922 *Z. tech. Phys.* **3**, 290–4, 320–7, 349–57.

— 1925 *Z. tech. Phys.* **6**, 166–72.

— 1927 *Z. tech. Phys.* **8**, 141–54.

— 1929 *Wiss. Veröff. Siemens-Konz.* **7**, 2, 217–71.

Holm 1931 *Wiss. Veröff. Siemens-Konz.* **10**, (4), 1–19, 20–64.
Holm, R. and Friman, E. 1926 *Z. tech. Phys.* **7**, 198–200.
Holm, R. and Holm, E. 1929 *Wiss. Veröff. Siemens-Konz.* **7**, 2, 272–305.
Knauer, F. and Stern, O. 1929 *Z. Phys.* **53**, 779–91.
Maxwell, C. 1873 "Electricity and Magnetism", **1**, art. 308.
Meyer, A. 1898 *Öfvers. VetenskAkad. Förh., Stockh.*, **55**, 199–215.
Nadai, A. 1931 "Plasticity". New York: McGraw-Hill Book Company.
O'Neill, H. and Greenwood, H. 1932 *J. Inst. Met.* **48**, 47–67.
Pedersen, P. O. 1916 *Electrician*, **76**, 589–91, 625–7.
Taylor, G. I. and Farren, W. S. 1925 *Proc. Roy. Soc.* A, **107**, 422–51.
Thomson, G. P. 1934 *Phil. Mag.* **18**, 640–56.

The electronic structure of some polyenes and aromatic molecules

VII. Bonds of fractional order by the molecular orbital method

By C. A. Coulson
University Chemical Laboratory and Trinity College, Cambridge[*]

(*Communicated by J. E. Lennard-Jones, F.R.S.—Received* 3 *September* 1938)

1. Introduction

It is known that in aromatic molecules the individual links have a length which is neither that of a pure single bond, nor of a pure double bond, but which is, in general, intermediate between the two. In earlier papers of this series (Lennard-Jones 1937; Penney 1937) two methods have been outlined for calculating these lengths and the energies associated with them. An approximate description of a particular link can be given if we assign to it an order. This order will be the sum of two parts, the one from the σ-bonds which is assumed to have the value unity in all cases, and the other from the π-bonds. We may call this latter contribution the "mobile order" p, since it arises from the motion of those electrons which Lennard-Jones (1937) calls "mobile electrons". For a pure single bond such as ethane, the total order is 1; for a pure double bond such as ethylene, the

[*] Now at University College, Dundee.

total order is 2; for a pure triple bond such as acetylene, the total order is 3. The corresponding values of p are 0, 1 and 2 respectively. For other links, as for example in benzene, p will have a fractional value.

Two theoretical calculations of the order of a bond have been given in the literature, both of which depend upon the valence-bond treatment of resonance, as developed by Pauling and others. In the first of these, due to Pauling, Brockway and Beach (1935), it is supposed that the wave function of the molecule in its ground state may be expanded in a series of canonical structures. Some of these structures will show a particular link as double, others as single, and the probability of a given structure is known from the coefficient of that structure in the wave function; the fractional order is then obtained by an appropriate summation. Penney (1937) showed that this definition was not quite satisfactory, because p could be changed by the addition of more excited canonical structures, so that, for example, in benzene, the addition of more structures increased the stability of the molecule, although it lowered the order of the bond. In addition, p depended upon the precise set of structures chosen as fundamental. An improvement upon Pauling's treatment could be obtained by use of the vector model of Dirac, defining the mobile order of a bond between atoms i and j to be

$$p_{ij} = -\tfrac{4}{3}\overline{s_i s_j},$$

where $\overline{s_i s_j}$ is the diagonal element of the product matrix $\mathbf{s}_i . \mathbf{s}_j$ for the state of least energy, and $\mathbf{s}_i$, $\mathbf{s}_j$ are the spin-vector matrices of the two-paired electrons from atoms i and j. By this means Penney was able to calculate the orders of the bonds of many of the simpler aromatic compounds.

Both the above methods become increasingly tedious as the compounds discussed grow more complex and the number of excited canonical structures rapidly multiplies. A new definition of the order of a bond is therefore proposed in the present paper, this time using the method of molecular orbitals. It will be found that the lengths predicted by this new method agree excellently with those predicted by Lennard-Jones and Penney, and also that the method can be applied in many cases without a great deal of labour.

2. PROPOSED ORDER OF A BOND, USING THE METHOD OF MOLECULAR ORBITALS

Let us use the notation of I, in which a given molecular orbital is expressed as the sum of atomic orbitals $\psi_1 \psi_2 \ldots \psi_k$ of the individual π-electrons. The molecular orbital will be

$$\Psi = a_1\psi_1 + a_2\psi_2 + \ldots + a_k\psi_k. \tag{1}$$

If Ψ is normalized, and if in addition we neglect the non-orthogonality of the ψ_i, then

$$\Sigma a_i^2 = 1. \tag{2}$$

We shall usually wish to consider a variety of different orbitals, so let us distinguish between them by using an affix n. The nth orbital is therefore

$$\Psi^{(n)} = a_1^{(n)}\psi_1 + a_2^{(n)}\psi_2 + \dots + a_k^{(n)}\psi_k, \tag{3}$$

and

$$[a_1^{(n)}]^2 + [a_2^{(n)}]^2 + \dots + [a_k^{(n)}]^2 = 1. \tag{4}$$

We may consider the mobile order of a given bond as the sum of separate contributions from each of the various π-electrons; it will be sufficient therefore to discuss the bonding of any one such electron; the total mobile order of a bond p_{ij} will be the sum of the partial mobile orders of the occupied molecular orbitals $p_{ij}^{(n)}$, and we may write

$$p_{ij} = \sum_{(n)} p_{ij}^{(n)}, \tag{5}$$

the summation being over the occupied orbitals n. We can easily state our definition of partial order with reference to wave function 3. Consider the partial bond between atoms i and j. Then unless i and j are neighbours in the usual sense, this partial bond order is zero; if they are neighbours, then the partial order is defined to be

$$p_{ij}^{(n)} = a_i^{(n)} a_j^{(n)}. \tag{6}$$

In the case in which the coefficients a are complex, this last definition has to be replaced by

$$p_{ij}^{(n)} = \tfrac{1}{2}[a_i^{(n)} a_j^{(n)*} + a_i^{(n)*} a_j^{(n)}]. \tag{7}$$

Thus the problem of determining the order of a given bond resolves itself into the calculation of the partial orders of the individual electrons; these partial orders are added together to get the total mobile order. A word must be said concerning the evaluation of the coefficients a_i. Strictly these can only be determined from the secular determinant when the lengths, and hence the resonance integrals β_{ij} (I, p. 285), are already known. But the calculations of I–VI show that as the molecules become more complex, the links tend to become more and more equivalent. Thus in butadiene (III, p. 320) the greatest difference between the links is 0·09 A, in hexatriene and octatetraene it is only 0·07 A. We may expect that in other more complex molecules the maximum difference will not exceed this value. As a result the resonance integrals β_{ij} are nearly equivalent, and no appreciable error will be introduced if the a_i are calculated on the supposition that all the β are equal. The advantage of this assumption is that then the a_i do not depend at all upon the actual numerical values of β. Our procedure may be called a first-order perturbation calculation of the type discussed by Lennard-Jones (I, p. 292). Since it is this step which makes the calculations considerably easier than the fuller ones of Lennard-Jones, Turkevich and Coulson (I, II, IV) it is desirable to justify it carefully; in a later section (p. 420) we shall give a justification in terms of the forces acting across the links; if further argument is desired, it may be found in the extraordinary agreement of our predicted bond lengths with those of Penney in III (see Table II).

The method will become clearer if we take an example; so to illustrate the definition (6) let us discuss the orders of the bonds in benzene. If we

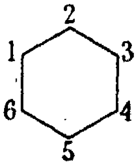


FIG. 1

assume that all the β are equal (fig. 1), then the occupied molecular orbitals of the π-electrons are

$$\left.\begin{aligned}
\Psi^{(1)} &= \frac{1}{6^{\frac{1}{2}}}[\psi_1+\psi_2+\psi_3+\psi_4+\psi_5+\psi_6], \quad \epsilon = -2\beta,\\
\Psi^{(2)} &= \frac{1}{6^{\frac{1}{2}}}[\psi_1+w\psi_2+w^2\psi_3+w^3\psi_4+w^4\psi_5+w^5\psi_6], \quad \epsilon = -\beta,\\
\Psi^{(3)} &= \frac{1}{6^{\frac{1}{2}}}[\psi_1+w^5\psi_2+w^4\psi_3+w^3\psi_4+w^2\psi_5+w\psi_6], \quad \epsilon = -\beta,\\
& \qquad w = \exp \pi i/3.
\end{aligned}\right\} \tag{8}$$

An electron in orbital $\Psi^{(1)}$ gives a partial bond order of $\frac{1}{6^{\frac{1}{2}}} \times \frac{1}{6^{\frac{1}{2}}} = \frac{1}{6}$ in each of the six bonds of the molecule. An electron in orbital $\Psi^{(2)}$ gives a bond between atoms 1 and 2 of value

$$\frac{1}{2}\left[\frac{1}{6^{\frac{1}{2}}} \times \frac{w^5}{6^{\frac{1}{2}}} + \frac{w}{6^{\frac{1}{2}}} \times \frac{1}{6^{\frac{1}{2}}}\right] = \frac{1}{12}.$$

The same bond strength is provided for the other bonds. Thus the total mobile order of the bonds in benzene is

$$(2 \times \tfrac{1}{6}) + (4 \times \tfrac{1}{12}) = \tfrac{2}{3}.$$

The first term is from the two electrons in orbital $\Psi^{(1)}$, and the second term is from the four electrons in orbitals $\Psi^{(2)}$ and $\Psi^{(3)}$. The final bond order, including the σ-bonds, is therefore 1·667.

Another illustration, which is interesting, is the case of cyclobutadiene (figs. 2A, 2B). As Wheland shows (V, p. 398), there are four possible orbitals:

$$\begin{aligned}
\Psi^{(1)} &= \tfrac{1}{2}(\psi_1+\psi_2+\psi_3+\psi_4), & \epsilon &= -2\beta, \\
\Psi^{(2)} &= \tfrac{1}{2}(\psi_1+\psi_2-\psi_3-\psi_4), & \epsilon &= 0, \\
\Psi^{(3)} &= \tfrac{1}{2}(\psi_1-\psi_2-\psi_3+\psi_4), & \epsilon &= 0, \\
\Psi^{(4)} &= \tfrac{1}{2}(\psi_1-\psi_2+\psi_3-\psi_4), & \epsilon &= +2\beta.
\end{aligned}$$

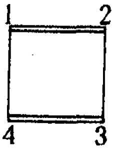


FIG. 2A

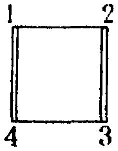


FIG. 2B

The partial orders of the bonds 1–2, 2–3, 3–4 and 4–1, in these four orbitals, are, respectively,

$$\begin{aligned}
&\Psi^{(1)} \ldots \quad \tfrac{1}{4}, \quad \tfrac{1}{4}, \quad \tfrac{1}{4}, \quad \tfrac{1}{4}, \\
&\Psi^{(2)} \ldots \quad \tfrac{1}{4}, -\tfrac{1}{4}, \quad \tfrac{1}{4}, -\tfrac{1}{4}, \\
&\Psi^{(3)} \ldots -\tfrac{1}{4}, \quad \tfrac{1}{4}, -\tfrac{1}{4}, \quad \tfrac{1}{4}, \\
&\Psi^{(4)} \ldots -\tfrac{1}{4}, -\tfrac{1}{4}, -\tfrac{1}{4}, -\tfrac{1}{4}.
\end{aligned}$$

Thus $\Psi^{(1)}$ gives equal bond strength for all bonds, $\Psi^{(2)}$ gives bonding between 1–2 and between 3–4, but antibonding between 2–3 and between 4–1. $\Psi^{(2)}$ clearly corresponds to the Kékulé-like structure 2*a* and $\Psi^{(3)}$ to the structure 2*b*. $\Psi^{(4)}$ is evidently a highly repulsive state in which there is repulsion in all bonds. There are four mobile electrons to allot to these orbitals. The first two will fill the lowest orbital $\Psi^{(1)}$, and the next two must be divided between $\Psi^{(2)}$ and $\Psi^{(3)}$. If we put both in $\Psi^{(2)}$ forming a singlet state, then we are automatically preferring the Kékulé structure 2*a*, and the total mobile orders of the four bonds are 1, 0, 1, 0 respectively; this shows that we have alternate pure-single and pure-double bonds, as Lennard-Jones and Turkevich found in II (p. 301). If we had put the last two electrons into $\Psi^{(3)}$ instead, we should merely have interchanged the single and double bonds. If one electron is in $\Psi^{(2)}$ and the other in $\Psi^{(3)}$, giving rise to a singlet and a triplet, we get total mobile orders $\tfrac{1}{2}$, $\tfrac{1}{2}$, $\tfrac{1}{2}$, $\tfrac{1}{2}$, so that all the links are equal. We have thus confirmed the work of Wheland in paper V, where this problem is discussed fully, and where it is shown that neither of the singlet states resulting from both electrons in $\Psi^{(2)}$ or both in $\Psi^{(3)}$ is satisfactory as a molecular wave function, but that they must be compounded together to give a singlet wave function of the correct symmetry; when this is done, it is soon seen that the links are all equal.

It is interesting to see what our definition of the order of a bond means in the simple diatomic case; thus, in the molecule H_2 each electron is described by an orbital

$$\frac{1}{\sqrt{2}}(\psi_a \pm \psi_b),$$

and the partial bond strength per electron is $\pm\frac{1}{2}$, agreeing with the usual description of the ground state of the H_2 bond as single. On this theory, the total order of the ground state of H_2^+ is $\frac{1}{2}$, and with an orbital

$$\frac{\lambda\psi_a + \mu\psi_b}{(\lambda^2+\mu^2)^{\frac{1}{2}}},$$

the partial order of the bond per electron is $\lambda\mu/(\lambda^2+\mu^2)$. This has a maximum value of $\frac{1}{2}$, when $\lambda = \mu$, so that the binding is homopolar; the order decreases with increasing polarity.

3. RELATION OF BOND ORDERS TO THE ENERGY

It is possible to relate the partial bond orders defined in the last section to the energies of the various orbitals. If we use a normalized wave function of type 3, then the energy of the orbital $E^{(n)}$ is

$$\begin{aligned} E^{(n)} &= (a_1^{(n)}\psi_1 + a_2^{(n)}\psi_2 + \dots + a_k^{(n)}\psi_k \mid H \mid a_1^{(n)}\psi_1 + a_2^{(n)}\psi_2 + \dots + a_k^{(n)}\psi_k) \\ &= \sum_{r=1}^{k} [a_r^{(n)}]^2 E_0 + 2\sum_{ij} a_i^{(n)} a_j^{(n)} \beta_{ij} \\ &= E_0 + 2\sum_{ij} a_i^{(n)} a_j^{(n)} \beta_{ij}. \end{aligned} \tag{9}$$

The notation is the same as that of paper I, so that

$$E_0 = \int \psi_r H \psi_r dv,$$

$$\beta_{ij} = \int \psi_i H \psi_j dv = \int \psi_j H \psi_i dv.$$

Let $\epsilon^{(n)}$ be the binding energy of the orbital $\Psi^{(n)}$, so that

$$\epsilon^{(n)} = E_0 - E^{(n)}. \tag{10}$$

Then, from (9) and (10), we have

$$\epsilon^{(n)} = -2\sum_{ij} p_{ij}^{(n)} \cdot \beta_{ij}. \tag{11}$$

Since this summation is taken over all ij, but by hypothesis all β_{ij} vanish except for nearest neighbours, this last summation is in effect a summation over the various bonds. When calculating the molecular-orbitals $\Psi^{(n)}$ we supposed that all the non-zero β_{ij} were equal ($= \beta$, say), and therefore, in the case of any single electron, we have a simple relation between the sum of its partial bonds and its binding energy

$$\sum_{ij} p_{ij}^{(n)} = -\epsilon^{(n)}/2\beta. \tag{12}$$

Quite properly, therefore, the greatest binding energy is associated with the greatest partial bond order. This last relation is very satisfactory; it shows that the total bonding power of an electron is directly proportional to its binding energy, and relates our proposed definition of order to that used by Penney, which again only depends on the binding energy. In fact, a comparison of the orders of the various bonds of a molecule, calculated according to the two definitions, becomes now a comparison of the two energies which the methods predict. The conception of partial bond strengths may prove useful in discussing the excitation of molecules, since the bond strengths of the excited electron can be calculated before and after excitation, and it can be seen which bonds are most weakened by the excitation. So also the most accessible positions of attack and substitution by other radicals may be found, along lines similar to those laid down by Pauling and Wheland (1935).

By summation of equation (12) over all the occupied orbitals, we obtain a relation between the total mobile orders and the total binding energy

$$\sum_{ij} p_{ij} = -\mathscr{E}/2\beta, \tag{13}$$

where $\mathscr{E} = \sum_{n} \epsilon^{(n)} =$ the total mobile binding energy.

The two equations (12) and (13) provide useful checks in the calculations.

An interesting application of this last equation occurs when all the bonds are equivalent, as in benzene, cyclobutadiene or graphite. In this case, if there are n bonds, each of mobile order p, then

$$p = -\mathscr{E}/2n\beta. \tag{14}$$

Thus for benzene (p. 416), where the occupied levels are $\Psi^{(1)}$, $\Psi^{(2)}$ and $\Psi^{(3)}$ (each occupied twice), and

$$\mathscr{E} = -8\beta, \quad n = 6,$$

we have

$$p = 8\beta/12\beta = +\tfrac{2}{3}.$$

This is precisely the result obtained before by use of the detailed wave functions but by more elaborate methods.

We can now give the physical interpretation of the bond order in terms of forces, which we referred to on p. 416. Suppose we sum equation (11) over all the occupied orbitals:

$$\mathscr{E} = -2\sum_{ij} p_{ij}\beta_{ij}. \tag{15}$$

It follows therefore that

$$p_{ij} = -\frac{1}{2}\frac{\partial\mathscr{E}}{\partial\beta_{ij}}. \tag{16}$$

For the purposes of this partial differentiation, all the β_{ij} are first put equal and then a small change is made in that particular β_{ij} with which we are concerned. Equation (16) shows us that p_{ij} is a measure of the change of energy when β_{ij} (and therefore the length of the link x_{ij}) is altered. In other words, in our proposed definition, the order of a given bond is a measure of the force tending to change the length of that bond, in the configuration in which all the bonds are supposed equal in length, only forces due to the mobile electrons being considered. This interpretation of the bond order may be said to justify our proposed definition.

4. THE BOND ORDERS OF THE CONJUGATED CHAINS $C_{2n}H_{2n+2}$

We may apply the analysis of the last section to a discussion of the general conjugated chain molecule $C_{2n}H_{2n+2}$, in which there are $2n-1$ links, whose lengths, in order, may be taken to be $x_1, x_2, \ldots, x_{2n-1}$. By symmetry $x_r = x_{2n-r}$. Our problem will be to calculate the order of the link x_r; in § 6 we shall use this order to deduce the length. The secular determinant, without simplification, is

$$\Delta \equiv \begin{vmatrix} \epsilon, & \beta_1, & 0, & \cdots & \cdots & & & 0, \\ \beta_1, & \epsilon, & \beta_2, & \cdots & \cdots & & & 0, \\ & & & \cdots & \cdots & & & \\ 0, & & & \cdots & \cdots & \beta_{2n-2}, & \epsilon, & \beta_{2n-1}, \\ 0, & & & \cdots & \cdots & 0, & \beta_{2n-1}, & \epsilon, \end{vmatrix} = 0.$$

Δ contains $2n$ rows and columns, and β_r is the exchange integral across the link x_r. First suppose that all the links are equal, and therefore put every $\beta_r = \beta_0$. In the appendix to paper IV of this series (equation 34) it is shown that the determinant then becomes

$$\begin{aligned}\Delta_0 &= P_{2n}(\beta_0) \\ &= \beta_0^{2n} \sin(2n+1)\,\theta/\sin\theta, \qquad (17)\end{aligned}$$

in which, as usual, we have written

$$\epsilon = 2\beta_0 \cos\theta. \qquad (18)$$

The roots of the determinant are

$$\epsilon^{(j)} = 2\beta_0 \cos\theta_j,$$

where
$$\theta_j = j\pi/(2n+1), \quad j = 1, 2, \ldots, 2n. \qquad (19)$$

Now suppose that across the link x_r we make a small change in β_r and put $\beta_r = \beta_0 + b_r$, all the other exchange integrals retaining the value β_0. The secular determinant is now

$$\Delta \equiv \begin{vmatrix} \epsilon, & \beta_0, & 0, & \ldots & & \ldots & & & & 0, \\ \beta_0, & \epsilon, & \beta_0, & \ldots & & \ldots & & & & 0, \\ 0, & \beta_0, & \epsilon, & \ldots & & \ldots & & & & 0, \\ & & \ldots & \beta_0, & \epsilon, & \beta_0+b_r, & 0, & \ldots & & \\ & & \ldots & 0, & \beta_0+b_r, & \epsilon, & \beta_0, & \ldots & & \\ 0, & & \ldots & & & & & & \epsilon, & \beta_0, \\ 0, & & \ldots & & & & & & \beta_0, & \epsilon, \end{vmatrix} = 0.$$

This may be expanded in powers of b_r, and we find that

$$\Delta = P_{2n}(\beta_0) - 2b_r\beta_0 P_{r-1}(\beta_0)\,P_{2n-r-1}(\beta_0) - b_r^2 P_{r-1}(\beta_0)\,P_{2n-r-1}(\beta_0).$$

The roots of this equation may be found as a series of approximations, in powers of b_r, of which the first is given by equation (19). The second approximation is

$$\epsilon^{(j)} = 2\beta_0 \cos j\pi/(2n+1) + \frac{4b_r}{2n+1} \sin jr\pi/(2n+1) \sin(r+1)\,j\pi/(2n+1).$$

The occupied levels are those with $j = 2n, 2n-1, \ldots, n+1$, and each level is occupied twice. The total electronic energy of the mobile electrons is therefore

$$\mathscr{E} = -2 \sum_{j=n+1}^{2n} \epsilon^{(j)}.$$

We can now evaluate $\frac{\partial \mathscr{E}}{\partial \beta_r}$, since it is just the coefficient of b_r in this last equation. Therefore

$$p_r = -\frac{4}{2n+1}\sum_{j=n+1}^{2n} \sin\frac{rj\pi}{2n+1}\sin\frac{(r+1)j\pi}{2n+1}$$

$$= \frac{1}{2n+1}\{\operatorname{cosec}\pi/(4n+2) + (-1)^{r-1}\operatorname{cosec}(2r+1)\pi/(4n+2)\}. \quad (20)$$

This formula gives the total mobile order of the rth link in the general conjugated chain $C_{2n}H_{2n+2}$. Table I gives the calculations performed with this formula. In each case the links are numbered from one end, and the last figure on each line therefore refers to the central link for that molecule. The results found by Penney (III) are shown for comparison, and it will be seen that the differences are not very great.

TABLE I. ORDERS OF LINKS IN THE CONJUGATED CHAINS $C_{2n}H_{2n+2}$

		End link x_1		x_2		x_3		x_4	
Molecule	n	Coulson	Penney	Coulson	Penney	Coulson	Penney	Coulson	Penney
Ethylene	1	2·00	2·00	—	—	—	—	—	—
Butadiene	2	1·894	1·911	1·447	1·333	—	—	—	—
Hexatriene	3	1·871	1·883	1·483	1·374	1·785	1·806	—	—
Octatetraene	4	1·862	1·88	1·495	1·37	1·758	1·76	1·529	1·46

5. RELATION BETWEEN ORDER AND LENGTH

We may now assume that the fractional order of any link is known; examples of this calculation for more complex molecules will be given in later papers by the author. The next problem is to use the calculated orders to deduce the lengths of the various bonds. For this purpose a curve relating the order and the length, similar to that used by Penney (III, p. 317) is necessary. Certain points on this curve are soon obtained. Thus ethane, with order 1 and length 1·54 A, ethylene, with order 2 and length 1·33 A, and acetylene, with order 3 and length 1·20 A are already known (see III, pp. 317–18). In benzene the observed length is 1·39 A, and we have already calculated the order to be 1·667. Another suitable point on the order-length curve is provided by graphite. For this crystal, Wheland (1934) has calculated an energy of $1{\cdot}58\beta$ per C atom; but n atoms give rise to $3n/2$ bonds, so that the energy per bond is $1{\cdot}06\beta$. It follows from equation (14) that the mobile order is 0·53 and therefore the total order is 1·53. These five points enable a curve to be drawn relating the bond order and the length of link. It is seen (fig. 3) that the points do lie on a very smooth curve, and this curve may now be used to predict the lengths of other links.

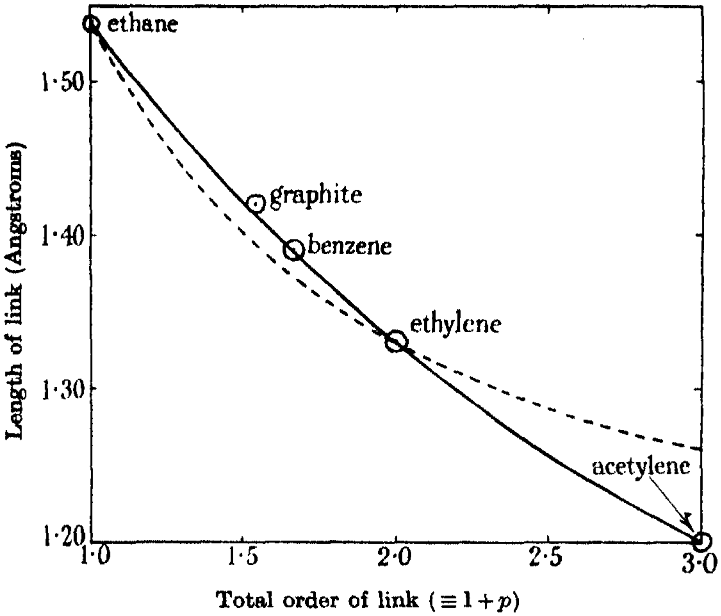


FIG. 3

However, before we actually do calculate any lengths in particular cases, it is instructive to relate our work to the ideas of Lennard-Jones in paper I. Let us suppose that equation (15), calculated on the assumption that all the β_{ij} are equal, also holds for other values of β_{ij}. (This is equivalent to using first-order perturbation theory, and must ultimately be justified a posteriori on its results.) Then the total energy $\mathscr{F}$ (I, p. 283) which allows for the compression of the σ-bonds, is

$$\mathscr{F} = \sum_{ij} k_s(x_{ij}-s)^2 + 2\sum_{ij} p_{ij}\beta_{ij}. \qquad (21)$$

The first term is the compression energy of the link x_{ij} and the second is the energy of the π-electrons; the summation is over all the bonds. Each x_{ij} has to be chosen so that $\mathscr{F}$ is minimized, i.e.

$$\frac{\partial \mathscr{F}}{\partial x_{ij}} = 0 \quad \text{for all } i, j.$$

This gives
$$k_s(x_{ij}-s) + p_{ij}\frac{\partial \beta_{ij}}{\partial x_{ij}} = 0.$$

Lennard-Jones has shown (I, equation (41)) that since 2β is the mobile energy of the ethylene molecule

$$2\beta = E_d^0 - E_s^0 + k_d(x-d)^2 - k_s(x-s)^2. \tag{22}$$

The last two equations now give, for the length of a bond whose mobile order is p,

$$x = s - \frac{s-d}{1+\dfrac{k_s}{k_d}\left(\dfrac{1-p}{p}\right)}. \tag{23}$$

This equation may be used to predict the length of any link, once its mobile order p has been calculated. There is, however, some doubt as to the proper value of k_s/k_d to be used. Lennard-Jones (I, p. 294), on the basis of vibration constants for ethane and ethylene, calculated by Dennison and Sutherland, assumes that

$$k_s = 2{\cdot}48 \times 10^5, \quad k_d = 4{\cdot}90 \times 10^5 \text{ dynes/cm.},$$

so that $k_s/k_d = 0{\cdot}507$. The curve which results from this value of k_s/k_d is shown dotted in fig. 3. It will be seen that although the two curves predict almost equivalent lengths in the range between single and double bonds, there is considerable error in the region of the triple bond. This error can be eliminated by a different choice of k_s/k_d. If for the moment we regard this ratio as a quantity to be determined from the experimental order-length curve rather than the Morse curves of ethane and ethylene, then we shall choose k_s/k_d so that the triple bond of acetylene ($p = 2$) satisfies equation (23). This means that k_s/k_d must be given the value 0·765. With this value, the theoretical curve given by equation (23) and the experimental order-length curve (the full line of fig. 3) become identical, the differences nowhere exceeding 1/200 A. Since this difference is well within the present observational limits of error, it will be quite good enough to take this value for the ratio k_s/k_d, and the predicted lengths given in the following section have been obtained from equation (23), using this value.

The question now arises, can this change of k_s/k_d from 0·507 to 0·765 be justified, other than by the empirical fact that it predicts the right lengths of links? The writer believes that it can. We have to bear in mind that (1) the experimental uncertainties in k_s and k_d are fairly large, because of the interpretation of the spectrum, and, in addition, their precise values depend somewhat on the type of force field used; (2) 0·507 is obtained by a discussion of the links of ethane and ethylene; it does not follow that their values will be unaltered when we consider more complicated molecules where "end-effects", if any, may be expected to be less important. Thus, in octatetraene, for example, there are two links at the end of the chain, and five internal ones, for which k_s/k_d may possibly be different from the value for the end-links; in benzene there are no end-links at all. The work of Sklar (1937) shows that in ethylene there is a considerable percentage of ionic character; this effect will make a contribution to k_d which will be different for inner and end-links; (3) k_s and k_d are the values of the curvature at the bottom of the two Morse curves; they represent the parabolas which best approximate to the Morse curves. A Morse curve, however, as fig. 4 shows, is not at all

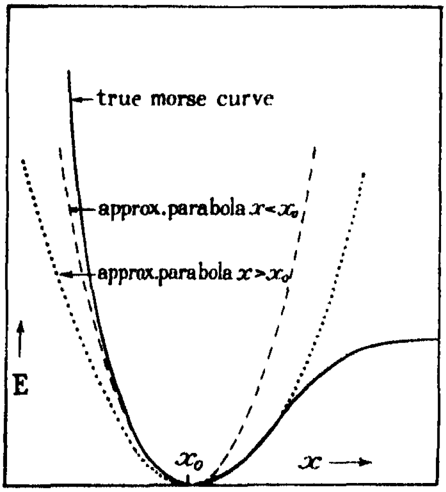


FIG. 4

symmetrical about its minimum, and we ought really to represent it by two parabolas, shown by dots and dashes in fig. 4; one of these parabolas would be relevant on the long, the other on the short, side of the true value. Since all the links, with which we are concerned, have lengths less than s, the appropriate value of k_s should be increased, perhaps quite considerably. Between single and double bonds k_d should be decreased, and between double and triple bonds k_d should be increased. It is difficult to assess exactly the magnitudes of these changes, but it is clear that in the significant part of the range the ratio k_s/k_d should be increased. Taking into consideration all three points above, there seems no valid reason why the value of k_s/k_d suited to our calculations should not be as high as 0·765. It must be remembered that this value for the ratio need not necessarily be the same as k_s/k_d for ethane and ethylene. The uncertainties of these latter calculations make it desirable that other data should be used to determine the appropriate value for our purposes, and the experimental length of acetylene is obviously the most satisfactory value to take. If we do so, then the calculations of order and length are carried through using only the experimental lengths of the single, double and triple bonds.

6. LENGTHS OF LINK IN CONJUGATED CHAINS

As an illustration of the methods outlined above, let us calculate the lengths of the links of the conjugated chains $C_{2n}H_{2n+2}$. The orders have already been obtained in § 4; it is an easy matter to use Table I and equation (23) to write down the lengths of the links. Table II shows the lengths thus determined, and again for comparison, the values given by Penney are included. It will be seen that the agreement is excellent throughout, the differences never exceeding 1/100 A. The extraordinary similarity between the values predicted by the two methods gives considerable reliability to the figures given. The lengths predicted by Lennard-Jones in paper I are also shown. It is evident that the hypothesis made in I that the links alternate in length is correct, but the further hypothesis that all the even-numbered links have one length and all the odd-numbered links another, is too general, for the variations in the lengths become inappreciable near the middle of the long chains.

TABLE II. LENGTHS OF LINKS IN CONJUGATED CHAINS $C_{2n}H_{2n+2}$

		End-link x_1			Link x_2		
Molecule	n	*C*	*P*	*L–J*	*C*	*P*	*L–J*
Ethylene	1	$1{\cdot}33_0$	$1{\cdot}33_0$	$1{\cdot}33_0$	—	—	—
Butadiene	2	$1{\cdot}34_7$	1·34	1·34	$1{\cdot}43_2$	1·43	1·43
Hexatriene	3	$1{\cdot}35_1$	1·35	$1{\cdot}34_8$	$1{\cdot}42_4$	1·42	$1{\cdot}40_4$
Octatetraene	4	$1{\cdot}35_3$	1·35	$1{\cdot}35_2$	$1{\cdot}42_2$	1·42	$1{\cdot}40_0$

		Link x_3			Link x_4		
Molecule	n	*C*	*P*	*L–J*	*C*	*P*	*L–J*
Ethylene	1	—	—	—	—	—	—
Butadiene	2	—	—	—	—	—	—
Hexatriene	3	$1{\cdot}36_6$	1·36	$1{\cdot}34_8$	—	—	—
Octatetraene	4	$1{\cdot}37_1$	1·37	$1{\cdot}35_2$	$1{\cdot}41_5$	1·41	$1{\cdot}40_0$

N.B. In calculating the lengths of the links in this and the following table, it has been assumed that $s = 1{\cdot}540$ A, and $d = 1{\cdot}330$ A exactly. This is almost certainly not true, and the last figure in each length, written as a small subscript, has no absolute significance; it is of interest, however, for comparing relative lengths of links.

An interesting deduction from equation (20) concerns the case of the infinite chain $C_{2n}H_{2n+2}$ when $n \to \infty$. The asymptotic form of equation (20) gives the mobile order of the jth link as

$$p_j = \frac{2}{\pi}(1 + (-1)^{j-1}/(2j+1)). \tag{24}$$

Table III shows the orders and the lengths of the first six links starting from one end. These results agree excellently with those obtained by Penney by the valence-bond method, and they give, as we go farther along the chain, an asymptotic order $1 + 2/\pi$, with a resulting length 1·394 A. This is almost the same as the value (1·38 A) found by Lennard-Jones (I, p. 294); it would have been exactly the same if we had used his value for k_s/k_d. The value for the end-link agrees quite well with that found by the author (1·34 A) in paper IV. There appear to be significant variations in the lengths of the links up to the seventh from each end.

TABLE III. ORDERS AND LENGTHS OF LINKS IN THE INFINITE CONJUGATED CHAIN

Link no.	1	2	3	4	5	6	Asymptotic value
Order	1·849	1·509	1·727	1·566	1·694	1·587	1·637
Length	$1{\cdot}36_5$	$1{\cdot}41_9$	$1{\cdot}37_7$	$1{\cdot}41_4$	$1{\cdot}38_3$	$1{\cdot}40_4$	$1{\cdot}39_4$

7. CONCLUSION

The methods outlined in this paper may be applied to more complex aromatic compounds; in a following paper the writer has applied them in a detailed study of naphthalene and the various hydro-naphthalenes; other applications will be published in a further paper. In every case the method predicts orders and lengths, which, though they cannot yet be checked in detail by experimental methods, are in general agreement with the bond diagrams of the chemist.

The writer gratefully acknowledges his indebtedness to Professor Lennard-Jones for advice and criticism in the preparation of this paper.

SUMMARY

A new definition of the order of a given bond in an aromatic compound is proposed, based upon the method of molecular orbitals. This order depends upon the coefficients of the wave functions in the occupied molecular orbitals; it is also related to the forces acting along the given link. A curve is drawn relating the order of a bond and its length, and the method is applied to a discussion of the orders and lengths in the conjugated chain molecules $C_{2n}H_{2n+2}$.

References

Coulson 1938 *Proc. Roy. Soc.* A, **164**, 383, IV.
Lennard-Jones 1937 *Proc. Roy. Soc.* A, **158**, 280, I.
Lennard-Jones and Turkevich 1937 *Proc. Roy. Soc.* A, **158**, 297, II.
Pauling, Brockway and Beach 1935 *J. Amer. Chem. Soc.* **57**, 2705.
Pauling and Wheland 1935 *J. Amer. Chem. Soc.* **57**, 2086.
Penney 1937 *Proc. Roy. Soc.* A, **158**, 306, III.
Penney and Kynch 1938 *Proc. Roy. Soc.* A, **164**, 409, VI.
Sklar 1937 *J. Chem. Phys.* **5**, 669.
Wheland 1934 *J. Chem. Phys.* **2**, 477.
— 1938 *Proc. Roy. Soc.* A, **164**, 397, V.

The spectrum of mono-deuteromethane in the photographic infra-red

I. Preparation of the gas, and measurements of a simple band at 9021 cm.$^{-1}$

By W. H. J. Childs, D.Sc., and H. A. Jahn, Ph.D.

(*Communicated by Sir William Bragg, O.M., P.R.S.—Received* 17 *October* 1938)

[Plate 27]

Introduction

An investigation of the infra-red spectra of the deutero substituted methanes is necessary for the complete spectroscopic determination of the structure of the methane molecule, for it was early recognized that for highly symmetrical molecules the normal frequencies are insufficient to determine uniquely the molecular parameters. In the case of methane the rotation structure is much influenced by interaction phenomena, and here again the spectra of the deuteromethanes can provide valuable material for the elucidation of the effects. We have recently completed the examination of the absorption spectrum of mono-deuteromethane in the photographic infra-red, from 8400 to 13,000 cm.$^{-1}$ (7600–12,000 A), and we propose here to give an account of the preparation of the gas, of which we required the relatively large amount of 40 l., and to present the measurements of a band at 9021 cm.$^{-1}$ having a simple rotation structure.

PREPARATION OF THE GAS

A number of preliminary experiments convinced us that the most suitable method was the Grignard synthesis:

$$CH_3MgI + D_2O = CH_3D + MgIOD,$$

with the more usual solvent diethyl ether replaced by the less volatile diamyl ether. Since we needed 40 l. of gas the yield of the reaction was important. If water is added directly to the ethereal solution of Grignard compound a yield of only 20 % is obtained. We suspected that the reason for this was that the lumpy precipitate of MgIOD trapped much of the water and withheld it from the reaction, and as heating and stirring failed to improve matters we added dioxane as an inert diluent to slow down the reaction and disperse the precipitate. When dioxane is added in excess to the ethereal solution of Grignard compound a precipitate is formed which, however, on shaking forms a smooth cream. When water diluted with dioxane is added to this there is a steady evolution of methane and, moreover, the mixture retains its creamy consistency. With this modification we were able, without heating the reagents or waiting for longer than a few minutes after all the water had been added, to obtain a yield of 93 % with small quantities and 70 % in the main experiment.

A number of preliminary experiments were also made to verify the purity of the gas involved in the reaction. Methane is difficult to purify, but we assumed that the following procedure would eliminate everything but the several methanes. First the gas was distilled from liquid air. The product was then reliquefied, and the pressure over the liquid reduced to a millimetre or so. The liquid solidified to a white crystalline mass which broke violently into small pieces with loud cracking noises. Pumping was continued for about 1 min., when the pump was shut off and the liquid air removed. The density of the gas obtained in this way from ordinary water was 0·720 ± 0·001 g./l. N.T.P. Cylinder methane which had been passed through a mixture of calcium chloride and soda lime also gave exactly the same value.

The density of the gas obtained from 99 % heavy water was 0·763 ± 0·001 g./l. N.T.P. Since the standard values are respectively 0·7167 and 0·7614, we felt justified in believing that in the second case our product was at least 90 % CH_3D. The correct density could of course be given by a mixture of ordinary methane and the heavier methanes in suitable proportions, but if (as subsequently proved to be the case) the gas showed not even the strongest lines of the spectrum of ordinary methane we could then be sure that the heavier methanes were also absent.

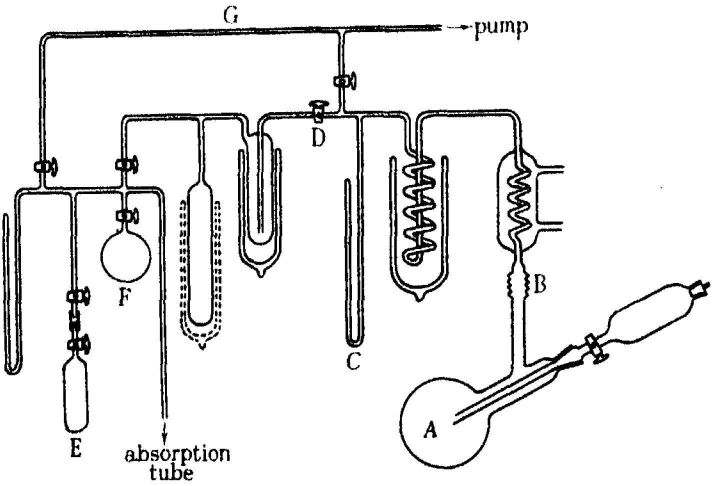


Fig. 1. Apparatus used for the preparation of CH_3D.

For the main synthesis it was decided to use 50 g. of 99·6 % heavy water, with the Grignard reagent somewhat in excess. The methyl iodide and amyl ether for the latter were redistilled and dried over calcium chloride and sodium respectively, whilst the dioxane was both redistilled and crystallized, and then dried over sodium. The reagent was prepared in three equal lots. For each were required (*a*) 62 c.c. methyl iodide mixed with 62 c.c. amyl ether, (*b*) 24·5 g. magnesium turnings, (*c*) 345 c.c. amyl ether. To start the reaction 30–40 c.c. of (*a*) were added to the magnesium. Then the amyl ether (*c*) was slowly added in two or three lots. When the reaction was apparently ended the rest of (*a*) was added. The total volume of ethereal solution of Grignard reagent made in this way was 1355 c.c. To generate the gas the glass apparatus shown in the accompanying diagram was constructed. The round-bottomed flask *A* contained initially Grignard reagent to which was then added an equal volume of dioxane. This was frozen in ice and the apparatus exhausted as far as the first tap, the remainder of the apparatus having already been exhausted via the tube *G*. The first and second condensers were then cooled with ice-water and alcohol-CO_2 mixture respectively, and the flask allowed to warm up and its contents to melt. Heavy water in ten times its volume of dioxane was then slowly added from the dropping funnel and thoroughly mixed with the Grignard reagent, a limited movement of the flask being permitted by the metal flexible joint *B*. The gas evolved was allowed to collect in the flask until the manometer *C* showed a pressure slightly less than atmospheric. The tap *D* was then turned and that particular charge of gas was passed into the first liquid-air trap and liquefied. Tap *D* was then closed and the process repeated as long as gas was evolved. From the first trap the gas was distilled to a second. Both traps contained thermoelements so that the temperature during distillation could be watched. From the second trap the gas was distilled into the container, and samples could be drawn off into the weighing bottle *E* or into a magnetically controlled micro-balance *F*.* The latter was useful in that it permitted a continuous check on the density. The gas was generated in four lots, the total yield being almost exactly 40 l. N.T.P. which corresponds to a 72 % efficiency. A sample of the gas was drawn off and treated in the way already described. Its density was 0·762 ± 0·001 g./l. N.T.P.

THE MEASUREMENTS

The gas was initially examined with the technique described by Childs (1936) and plates were successfully obtained of the region 8400–13,000 cm.$^{-1}$. A preliminary account of the measurements was published by Childs and Jahn (1936). A second investigation of this region was then made with a new absorption tube 16 ft. 3 in. long in which the light was reflected to and fro six times, giving a total path length of 97½ ft. With this tube a number of faint bands which were just visible on the first series of plates were brought up to measurable strength, the band at 9021 cm.$^{-1}$ (already reported on by Childs and Jahn 1936) by reason of its simple structure standing out especially well (see fig. 2, Plate 27). The measurements of this band are presented in Table I.

---

* For full details of the construction of such a balance see Stock and Ritter (1926).

TABLE I. MEASUREMENTS OF THE BAND AT 9021 CM.$^{-1}$

The intensities are merely eye estimations, 0 indicating lines which are certainly present but which are difficult to measure. For the stronger lines the accuracy should be better than 0·05 cm.$^{-1}$, but for the lines marked 0 only 0·10 cm.$^{-1}$. A considerable number of much fainter lines are present as a background; the measurements of these are still less accurate, and it was felt that no useful purpose would be served by including them here. A blend with a line due to water vapour is indicated by +w.v.

Intensity	Frequency *in vacuo*	Intensity	Frequency *in vacuo*
0	9113·54	0	8996·33
0	08·84	0	94·68
0	06·58	1	94·26
1	00·32	0	93·36
1	9099·95	1	91·81
0	98·21	0	90·96
1	93·48	0	90·28
1	93·10	1	89·19
2	86·29	1	88·81
2	79·39	0	87·82
2	72·44	0	87·31
3	65·41	3	84·90
2	58·19	1	83·62
2	50·90	0	82·81
1	43·52	1	81·95
0	36·05	0	81·23
0	28·41	0	80·43
1	27·12	0	79·59
1	22·91	1	78·82
0	21·18	0	77·32 + w.v.
0	20·36	0	76·95
2	19·88	1	76·48
0	19·40	2	75·86
1	18·90	1	75·46
0	18·50	2	74·60
0	16·01	2	73·54
0	14·02	0	72·93
0	13·27	0	72·66
1	10·01	0	72·49
1	05·16	0	71·84
2	04·13 + w.v.	1	71·40
0	02·53	1	70·42
1	01·69 + w.v.	1	69·65
1	01·44 + w.v.	1	69·34
0	00·88	0	68·47
0	00·48	0	67·85
0	8998·92 + w.v.	5	66·98
0	98·50 + w.v.	5	66·04
0	96·77 + w.v.	1	65·30

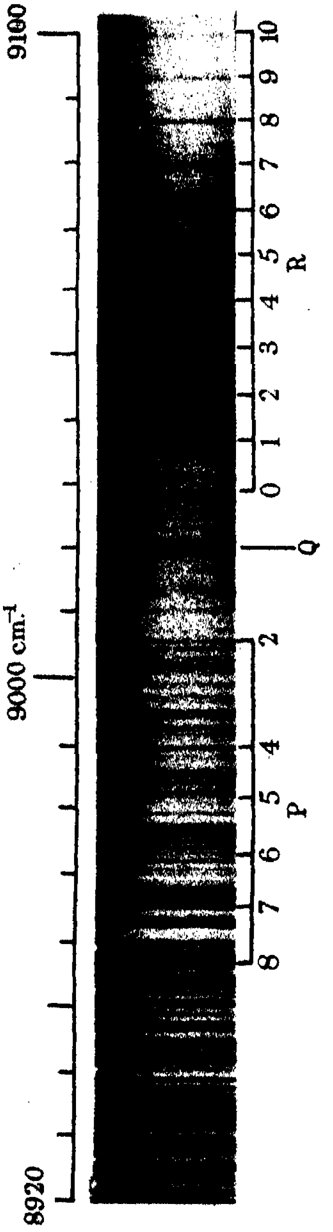


FIG. 2. An absorption band of CH_3D at 9021 cm.$^{-1}$.

TABLE I (*continued*)

Intensity	Frequency *in vacuo*	Intensity	Frequency *in vacuo*
2	8964·48 + w.v.	4	8944·02
0	63·28	0	43·14
2	62·64	1	42·36
2	61·69	5	41·55 + w.v.
0	60·55	0	40·94
2	59·09 + w.v.	0	40·42
0	58·42	1	39·78
1	57·76 + w.v.	1	39·57
3	57·09	4	38·33
4	56·12 + w.v.	4	37·17 + w.v.
2	55·67	0	36·22 + w.v.
0	55·07 + w.v.	2	35·35
3	54·08	1	34·65 + w.v.
1	53·42 + w.v.	2	34·28
1	52·76	2	33·53 + w.v.
0	52·24	2	32·93
0	51·97	2	32·51 + w.v.
0	51·45	1	31·92
0	51·18	3	31·16
2	50·29 + w.v.	0	30·25
3	49·73	3	29·48
0	48·97	2	28·45 + w.v.
1	48·31 + w.v.	2	27·67
3	46·88	2	27·34
2	46·00	0	26·64
0	44·90	10	25·28 + w.v.

DISCUSSION

The region covered by the measurements shows at the middle (see Plate 27) what appears to be a narrow Q branch, only partially resolved, with an R branch of approximately equidistantly spaced lines at its high-frequency side. Where the P branch should be there are many lines, most of which obviously belong to other bands. From its appearance, and the spacing of the lines (7·7 cm.$^{-1}$) in the R branch, this is clearly a parallel type band. The R, P and Q branch lines must therefore be given by expressions of the form

$$R_K(J-1) = \nu_0 + (B'' + B')\,J - (B'' - B')\,J^2 + [(B'' - B') - (A'' - A')]\,K^2,$$

$$P_K(J) \quad = \nu_0 - (B'' + B')\,J - (B'' - B')\,J^2 + [(B'' - B') - (A'' - A')]\,K^2,$$

$$Q_K(J) \quad = \nu_0 - (B'' - B')\,J(J+1) + [(B'' - B') - (A'' - A')]\,K^2,$$

where K is the axial, J the total rotational quantum number and

$$A = \frac{h}{8\pi^2 c I_A}, \quad B = \frac{h}{8\pi^2 c I_B}.$$

Here I_A = moment of inertia about the C-D axis (figure axis) and I_B = moment of inertia about an axis perpendicular to the figure axis. The symbols $''$ and $'$ are used to distinguish the ground and upper vibrational states respectively.

We find that the R branch is best fitted by

$$R(J-1) = 9020{\cdot}84 + 7{\cdot}707J - 0{\cdot}0481J^2, \tag{1}$$

and since the line width does not appreciably increase with increasing J value, the coefficient of the term in K^2, i.e. $[(B''-B')-(A''-A')]$, must be very small. A comparison between the frequencies observed and those computed from expression (1) is shown in Table II, cols. 2 and 3. The same table shows in cols. 5 and 8 the frequencies computed for the Q and P branch lines. Of the Q branch not very much can be said other than that it is certainly present and in its proper position, for unfortunately its fine structure is mostly just beyond the limit of the resolving power of the grating. A search for the P branch, however, shows that either it must be very much fainter than one would expect, with only a few lines in the region of the maximum showing up at all, or else it is strongly perturbed. Precisely this phenomenon is shown in an analogous band at 9047 cm.$^{-1}$ in the spectrum of ordinary methane.

TABLE II. COMPARISON OF CALCULATED AND OBSERVED FREQUENCIES

	R branch			*Q* branch			*P* branch	
	Calc.	Obs.		Calc.	Obs.		Calc.	Obs.
R (0)	9028·50	9028·41 (0)	*Q* (1)	9020·74	—	*P* (1)	9013·08	—
R (1)	9036·06	9036·05 (0)	*Q* (2)	9020·55	—	*P* (2)	9005·24	9005·16 (1)
R (2)	9043·53	9043·52 (1)	*Q* (3)	9020·26	9020·36 (0)	*P* (3)	8997·29	—
R (3)	9050·90	9050·90 (2)	*Q* (4)	9019·88	9019·88 (2)	*P* (4)	8989·24	8989·19 (1)
R (4)	9058·18	9058·19 (2)	*Q* (5)	9019·40	9019·40 (0)	*P* (5)	8981·10	8981·23 (0)
R (5)	9065·35	9065·41 (3)	*Q* (6)	9018·82	9018·90 (1)	*P* (6)	8972·87	8972·93 (0)
R (6)	9072·43	9072·44 (2)	*Q* (7)	9018·15	9018·50 (0)	*P* (7)	8964·53	8964·48 (2)
R (7)	9079·42	9079·39 (2)	*Q* (8)	9017·38	—	*P* (8)	8956·10	8956·12 (4)
R (8)	9086·30	9086·29 (2)	*Q* (9)	9016·51	—	*P* (9)	8947·58	—
R (9)	9093·10	9093·10 (1)	*Q* (10)	9015·55	—	*P* (10)	8938·96	—
		9093·48 (1)						
R (10)	9099·80	9099·95 (1)						
		9100·32 (1)						
R (11)	9106·41	9106·58 (0)						

From expression (1) we are able to calculate the value of the moment of inertia I_B. Since B'' is 3·878 cm.$^{-1}$, I_B comes out to be $7\cdot202\times10^{-40}$ g. cm.^{2}. To compare with this we have another value of I_B obtained from Ginsburg and Barker's (1935)* measurements on a band at 2205 cm.$^{-1}$. They give the expression

$$\nu = 2205\cdot25 \pm 7\cdot714J - 0\cdot0434J^2$$

as best representing their measurements, and from this in a similar way one obtains

$$B'' = 3\cdot879; \quad I_B = 7\cdot200\times10^{-40}\,\text{g. cm.}^2,$$

a very satisfactory agreement.

It is now possible to compare the dimensions of the molecule, calculated from this moment of inertia, with those of CH_4. If we assume regular tetrahedral structure in both cases we arrive at

$$\text{C-H distance of } CH_3D = 1\cdot093_0\,\text{A},$$

$$\text{C-H distance of } CH_4 = 1\cdot092_1\,\text{A}.$$

The major part of the difference, which is real, is probably due to the fact that no correction has been applied for the anharmonicity of the zero-point vibration.

Adopting a mean value, which we believe to be fairly reliable, of $1\cdot092_5$ for the C-H distance, we have calculated the moments of inertia of all the deuteromethanes. In Table III explicit expressions in terms of the masses and the C-H distance are given as well as the numerical values.

A comparison of the vibrational levels of CH_4 with those of CH_3D should enable us to form some estimation of the anharmonicity of the vibrations. Calculations to this end are in progress, but until these are complete the assignment of vibrational quantum numbers to the band can only be very tentative. It is, however, tempting to believe that it is a parallel component of $3\nu_2$, the second overtone of the perpendicular fundamental at 3031 cm.$^{-1}$. This assignment would preserve the analogy with the CH_4 band at 9047 cm.$^{-1}$, which may be the second overtone of the corresponding CH_4 fundamental at 3020 cm.$^{-1}$.

In conclusion, we wish to thank Sir William Bragg, O.M., P.R.S., and the Managers of the Royal Institution, for the facilities afforded at the Davy-Faraday Laboratory where this work was carried out.

---

* Ginsburg and Barker give erroneously $I_B = 7\cdot166\times10^{-40}$ g. cm.^{2} or $7\cdot241\times10^{-40}$, if more recent values of the fundamental constants are used.

TABLE III. MOMENTS OF INERTIA OF THE DEUTEROMETHANES

Molecule	Moment of inertia	Numerical value (observed values in parentheses) $\times 10^{-40}$ g. cm.^{2}
CH_4	$I=\frac{8m_1}{3}r^2$	5·322 (5·318)
CH_3D	$I_A=\frac{8m_1}{3}r^2$	5·322
	$I_B=I_C$ $=\frac{1}{3}\frac{4m_1(3m_1+5m_2)+m_0(5m_1+3m_2)}{m_0+3m_1+m_2}r^2$	7·196 (7·202)
CH_2D_2	$I_A=\frac{1}{3}\frac{4m_1(2m_1+6m_2)+m_0(6m_1+2m_2)}{m_0+2m_1+2m_2}r^2$	6·502
	$I_B=\frac{4(m_1+m_2)}{3}r^2$	7·978
	$I_C=\frac{1}{3}\frac{4m_2(2m_2+6m_1)+m_0(6m_2+2m_1)}{m_0+2m_1+2m_2}r^2$	9·158
CHD_3	$I_A=I_B$ $=\frac{1}{3}\frac{4m_2(3m_2+5m_1)+m_0(5m_2+3m_1)}{m_0+m_1+3m_2}r^2$	8·537
	$I_C=\frac{8m_2}{3}r^2$	10·634
CD_4	$I=\frac{8m_2}{3}r^2$	10·634 (10·7)*

m_1 = mass of hydrogen = 1·00782m.
m_2 = mass of deuterium = 2·01373m.
m_0 = mass of carbon = 12·0007m.
m = mass per unit atomic wt = $1\cdot659\times10^{-24}$ g.
r = C-H distance = 1·092$_5$ A.

* Nielsen and Nielsen, 1938, *Phys. Rev.* **54**, 118 (recalculated).

SUMMARY

The preparation of 40 l. of mono-deuteromethane (CH_3D) is described. The purity of the gas was tested by density measurements and by the complete absence of even the strongest absorption lines of ordinary methane on the photographic plates obtained. The absorption spectrum of the gas has been investigated from 8400 to 13,000 cm.$^{-1}$ in the photographic infra-red,

and measurements are given of a particularly simple band at 9021 cm.$^{-1}$. From this band the moment of inertia of the molecule perpendicular to the C-D axis can be deduced and a value of $7 \cdot 202 \times 10^{-40}$ g. cm.^{2} is obtained. This corresponds (assuming a regular tetrahedral structure) to a C-H distance of 1·093 A, to be compared with 1·092 A in the case of CH_4. The band is assigned tentatively as the parallel component of $3\nu_2$, the second overtone of the perpendicular fundamental at 3031 cm.$^{-1}$.

REFERENCES

Childs, W. H. J. 1936 *Proc. Roy. Soc.* A, **153**, 555.
Childs, W. H. J. and Jahn, H. A. 1936 *Nature, Lond.*, **138**, 285.
Ginsburg, N. and Barker, E. F. 1935 *J. Chem. Phys.* **3**, 668.
Stock, A. and Ritter, G. 1926 *Z. phys. Chem.* **119**, 333.

Electron transfer in mercury vapour

BY F. L. ARNOT, PH.D.
Lecturer in Natural Philosophy
AND MARJORIE B. M'EWEN, B.SC.
Demonstrator in Physics, The University of St Andrews

(*Communicated by H. S. Allen, F.R.S.—Received* 17 *October* 1938)

INTRODUCTION

During some experiments involving the acceleration and subsequent retardation of positive ions of mercury we found that the retarding potential had to be twice as great as the accelerating potential in order to stop completely the flow of ions. Investigation showed that this was due to some ions being accelerated as doubly charged Hg^{2+} ions and retarded as singly charged Hg^+ ions, the change of charge being due to the Hg^{2+} ion capturing an electron from a neutral atom with which it collided in the space between the accelerating and retarding fields. Since this process of electron transfer in mercury vapour had not been previously studied in detail we modified the apparatus to enable it to be investigated.

Electron transfer between *singly* charged ions and neutral atoms of the type $He^{+} + Ne \rightarrow He + Ne^{+}$ was first observed by Harnwell (1927) for various combinations of the rare gas atoms with each other and with hydrogen and nitrogen. Hogness and Lunn (1927) simultaneously discovered the effect with helium and argon ions in nitric oxide. The process has been studied in more detail by Kallmann and Rosen (1930), who measured the decrease in intensity of beams of singly charged nitrogen and argon ions due to electron transfer.

Electron transfer between *doubly* charged ions and neutral atoms has been extensively studied for α-particles (Rutherford, Chadwick and Ellis 1930), the effect being first noticed by Henderson (1922). Although Kallmann and Rosen observed the transformations $Ar^{2+} \rightarrow Ar^{+}$ in argon, and $Hg^{2+} \rightarrow Hg^{+}$ for Hg^{2+} ions in mercury vapour and nitrogen, the process was not investigated in any detail.

The present paper contains the results of our investigation of the process $Hg^{2+} + Hg \rightarrow Hg^{+} + Hg^{+}$ for Hg^{2+} ions having energies from 140 to 400 electron-volts. The cross-section for the process is determined over this energy range; it is shown that a head-on collision is required, and that one of the resulting Hg^{+} ions is excited by the process.

APPARATUS

A scale drawing of the apparatus and a wiring diagram is shown in fig. 1. All metal parts with the exception of the filament were constructed from out-gassed nickel sheet spot-welded where necessary. In fig. 1 *a* the cylinder C_1, which was 2 cm. in diameter and 2 cm in length, was divided into two parts by the nickel gauze G_1. The upper part contained the spiral filament F, the leads to which passed through two fine quartz tubes in the sides of the cylinder. One end of this cylinder was closed by the solid nickel cap spot-welded to the cylinder, and the other end was fitted with the gauze G_2.

The cylinder C_2, which was 1·5 cm. in diameter and 1·5 cm. in length, was closed at each end by the gauzes G_3. This was supported on, and insulated from, the cylindrical shield C_3 of the Faraday cylinder by means of a sindanyo ring. C_3 contained the gauze G_4 which admitted ions to the Faraday cylinder C_4 which was insulated from the shield C_3 by two sindanyo rings as shown. The lead to the Faraday cylinder was enclosed in a fine quartz tube. It is evident from the figure that the construction of the lower part of the apparatus was such that ions passing at any angle through G_3 would be received by the Faraday cylinder and recorded by the galvanometer M_1. It is also clear that ions could reach the Faraday cylinder only from the field-free cylinder C_2 after passing through G_3 and G_4.

The entire apparatus was carried on a 5 cm. ground-glass joint fitting into a Pyrex tube which was connected to a mercury diffusion pump backed by a Hyvac pump, and to a McLeod gauge. The joint was water-cooled to allow baking of the apparatus by an electric furnace. Apiezon grease of negligible vapour pressure was used on all joints and taps in the vacuum system. The pressure of mercury vapour in the apparatus was the saturation vapour pressure at room temperature, i.e. about 10^{-3} mm. The pressure of residual gas read on the McLeod gauge was less than 10^{-5} mm. of Hg for all runs.

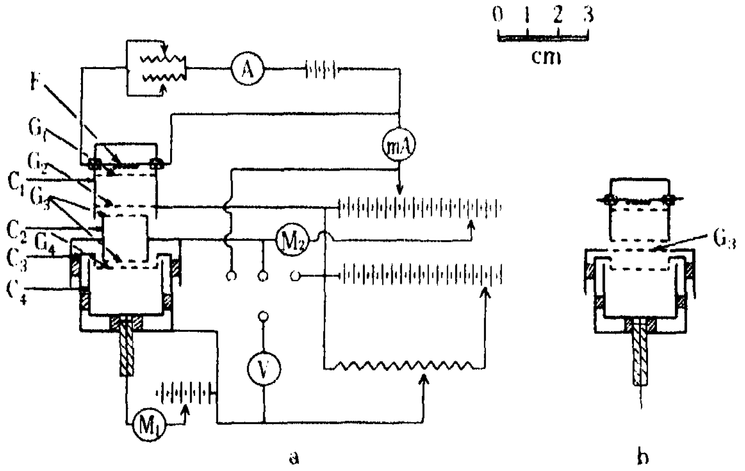


FIG. 1. Apparatus and wiring diagram.

The modified form of the apparatus shown in fig. 1*b*, obtained by substituting the single gauze G_3 in place of the cylindrical field-free cylinder C_2, was used to obtain the energy distribution of the positive ions passing through the gauze G_2.

EXPERIMENTAL PROCEDURE

The energy distribution of the positive ions that would enter the field-free cylinder C_2 in the main experiment was first obtained by using the modified form of the apparatus shown in fig. 1*b*. After the usual baking of the apparatus and the flashing and running in of the filament, the potential difference V_0 between the filament and G_1 was set at 60 volts. The positive ions formed in the field-free space between G_1 and G_2 were accelerated between G_2 and G_3 by a potential V_1, and retarded between G_3 and G_4 by a potential V_2. Those having sufficient energy to pass through G_4 were received by the Faraday cylinder and recorded by the galvanometer M_1. In order to prevent electrons from the filament reaching the Faraday cylinder the values given to V_1 were always greater than V_0, i.e. greater than 60 volts. All electrons from the filament were then reversed between G_2 and G_3 by the potential V_1. For a number of fixed values of the accelerating potential V_1 the retarding potential V_2 was varied in small steps from a value several volts below V_1 to a few volts above V_1. The curve representing the positive-ion current to the Faraday cylinder as a function of the retarding potential V_2 for each fixed value of V_1 was then differentiated to obtain the energy distribution of the ions for that value of V_1.

The apparatus was then changed to the form shown in fig. 1*a* by substituting the field-free cylinder C_2 in place of the single gauze G_3. For each of the above fixed values of V_1 a curve representing the positive-ion current to the Faraday cylinder as a function of the variable retarding potential V_2, now placed between G_3 and G_4, was then obtained and differentiated to give the energy distribution of the positive ions now reaching the Faraday cylinder.

For each value of V_1 we thus have curves representing (1) the energy distribution of the positive ions *entering* the field-free cylinder C_2 obtained with the apparatus of fig. 1*b*, and (2) the energy distribution of the positive ions *emerging* from the field-free cylinder C_2 obtained with the apparatus of fig. 1*a*. We shall call the positive ions entering C_2 the *incident ions*, and those passing out of C_2 through G_3 the *emergent ions*.

RESULTS

Fig. 2 shows the type of energy distribution curve obtained for the emergent ions. We find, as we would expect, a fairly homogeneous group of positive ions having an energy V_1 equal to the accelerating potential to which all the positive ions formed between G_1 and G_2 are subjected. The surprising and novel feature of the curve is the presence of a second group of ions having an energy of $2V_1$. The Hg^+ ions formed between G_1 and G_2 are given an energy of V_1 electron-volts between G_2 and G_3, and therefore cannot reach the Faraday cylinder when the retarding potential V_2 is greater than V_1. Similarly an Hg^{2+} ion accelerated between G_2 and G_3 by the potential V_1 would be prevented from reaching the Faraday cylinder by the same potential V_1 in a retarding direction between G_3 and G_4. But suppose the ion was accelerated between G_2 and G_3 as an Hg^{2+} ion and retarded between G_3 and G_4 as an Hg^+ ion, then it would require a retarding potential of $2V_1$ to stop it. The loss of one of the positive charges of the ion could be brought about by an electron transfer during a collision with a neutral atom in the field-free cylinder C_2. This is the explanation of the peak at $2V_1$ in the energy distribution curves of the emergent ions.

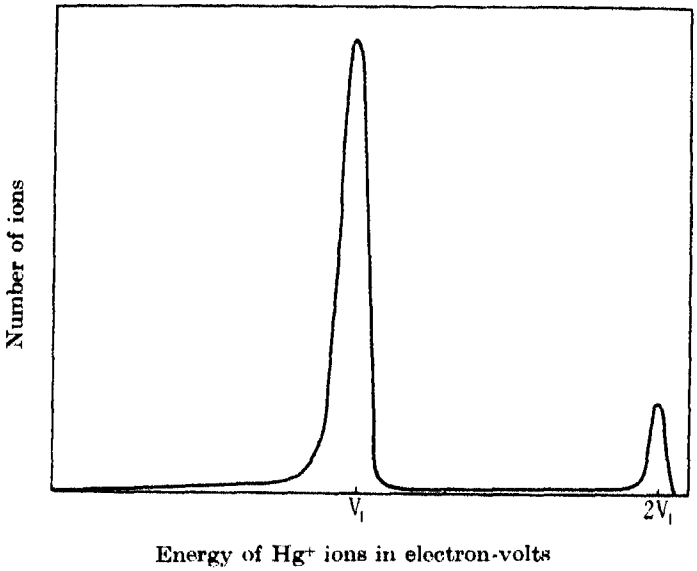


FIG. 2. Type of energy distribution curve obtained for the emergent ions. V_1 is the potential difference applied to accelerate the incident ions.

After acceleration by the potential V_1 the Hg^{2+} ion has an energy of $2V_1$ electron-volts. If this doubly charged ion now makes a collision with a neutral atom in C_2 *without loss of kinetic energy* but with an electron transfer between the colliding particles we have

$$Hg^{2+}(2V_1\,\text{electron-volts}) + Hg \rightarrow Hg^{+}(2V_1\,\text{electron-volts}) + Hg^{+} + p.$$

The extra energy p electron-volts gained in the collision is the difference between the ionization potential of the Hg^{2+} ion and the sum of the ionization potentials of the two resulting Hg^+ ions. The first ionization potential of mercury is 10·39 electron-volts, and the energy required to remove two electrons is 29·06 electron-volts (Bacher and Goudsmit 1932). Thus $p = 29{\cdot}06 - 270{\cdot}8 = 8{\cdot}28$ electron-volts.

If neither of the two Hg^+ ions is excited in the process this internal energy must be shared as kinetic energy between the two Hg^+ ions in accordance with the law of conservation of momentum. The mechanics of the collision process will be discussed later in connexion with the experimental results (see p. 446). At this stage it is only necessary to remark that p is very small compared to $2V_1$ which was always greater than 140 electron-volts, and therefore, neglecting p, the retarding potential V_2 between G_3 and G_4 required to stop this ion is $2V_1$.

If this is the correct explanation of the peak at $2V_1$ in fig. 2 it is clear that this peak should be absent if V_0 is reduced below the value necessary to form Hg^{2+} ions, i.e. below 29 volts. Fig. 3 shows the number of fast Hg^+ ions as a function of V_0, and it will be seen that the peak does not appear until V_0 is greater than 29 volts for each value of V_1.

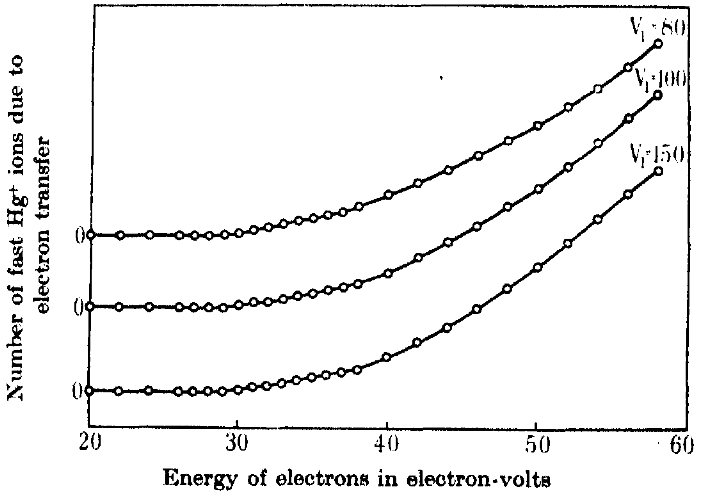


FIG. 3. Showing that fast Hg^+ ions due to electron transfer are not produced until the energy of the electron stream is greater than the critical potential for the formation of Hg^{2+} ions, 29 volts.

Fig. 2 is merely a sketch of the type of results obtained. The actual results for four different values of the accelerating potential V_1 are given in fig. 4. For each curve V_0 was 60 volts, which is below the critical potential, 71 volts (Bleakney 1930), for the formation of Hg^{3+}, thus ensuring that only Hg^+ and Hg^{2+} ions were formed between G_1 and G_2. Since we are concerned only with the fast ions resulting from the transformation $Hg^{2+} + Hg \rightarrow Hg^+ + Hg^+$ in the field-free cylinder C_2, the energy distribution curves were measured only for retarding potentials greater than V_1. The abscissae of these curves represent the retarding potential between G_3 and G_4, and thus give the energy in electron-volts of the fast Hg^+ ions resulting from electron transfer. Each energy distribution curve has its main peak occurring at a value equal to V_1 electron-volts and a second peak at a value equal to $2V_1$ electron-volts. For example, the accelerating potential between G_2 and G_3 was $V_1 = 70$ volts for the bottom curve of fig. 4. The curve shows one side of the large peak of normally accelerated and retarded Hg^+ ions at 70 electron-volts, and a small peak of ions resulting from electron transfer in C_2 at 140 electron-volts. The points on the second peak will be referred to later.

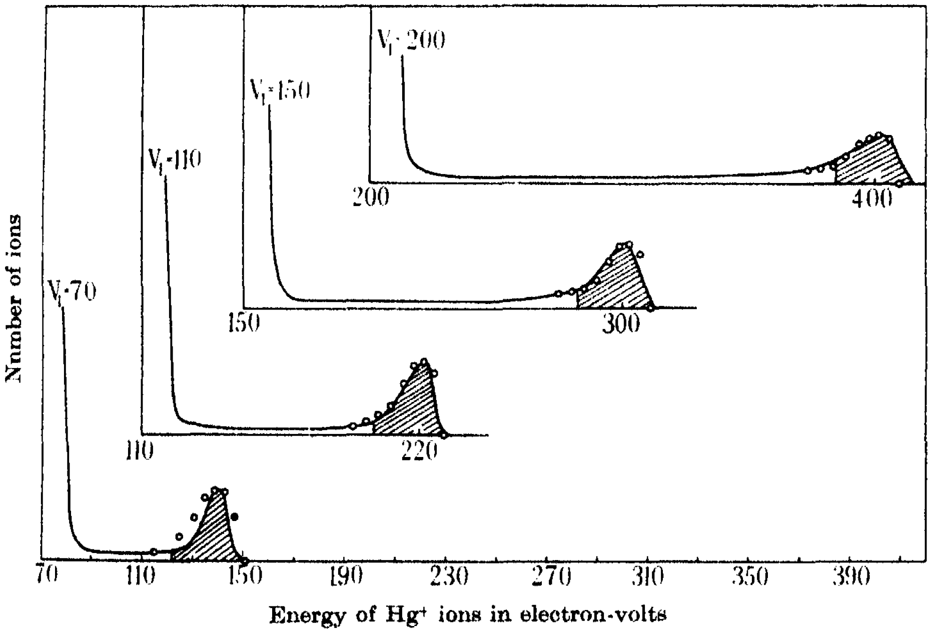


FIG. 4. Experimental energy distribution curves obtained for the emergent ions. V_1 is the potential difference applied to accelerate the incident ions. The points over the peaks are taken from the curves of fig. 5, and show that the energy spread of the emergent ions is entirely accounted for by the energy spread of the incident ions.

We have seen that the fast ions depicted in fig. 4 are due to a collision of a doubly charged ion in the field-free cylinder C_2 with a neutral atom in which an electron is transferred from the neutral atom to the ion, leaving both particles singly charged. Thus

$$Hg^{2+} + Hg \rightarrow Hg^+ + Hg^+.$$

We might expect the kinetic energy of the Hg^{2+} ion to be shared in a random way between the two Hg^+ ions. If this were so, we should not obtain a peak at $2V_1$ but, instead, the ions should appear over a wide range of energy between V_1 and $2V_1$. A sharp peak at a value $2V_1$ would mean that the above process is far more probable for either a head-on collision or a collision in which the Hg^{2+} ion just grazes the Hg atom. For both of these two types of collision one of the Hg^+ ions would carry off all the kinetic energy of the incident Hg^{2+} ion.

It is therefore of considerable importance to find whether the energy spread of the peaks in fig. 4 can be fully accounted for by an energy spread in the incident doubly charged ions. If it can be so accounted for, then the above process is far more probable for either a head-on or a grazing collision than for a collision in which one particle does not carry off all the energy. On the other hand, if the energy spread in the peaks of fig. 4 is definitely greater than that in the incident Hg^{2+} ions, then it would be possible to determine from this spread the angular distribution of the two Hg^+ ions resulting from the collision.

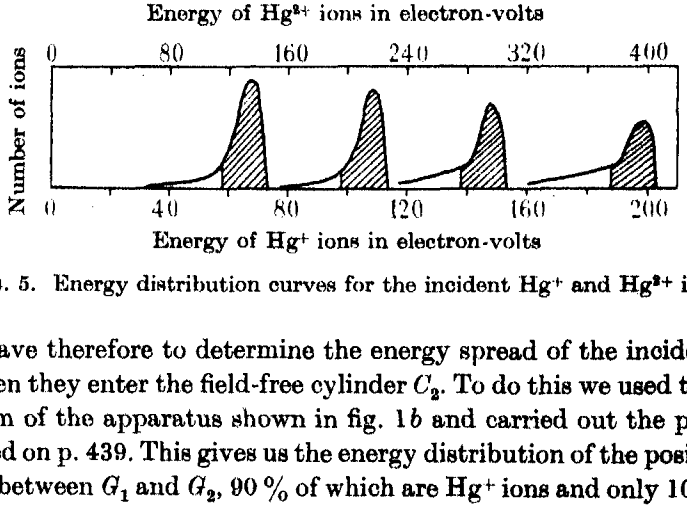


FIG. 5. Energy distribution curves for the incident Hg^+ and Hg^{2+} ions.

We have therefore to determine the energy spread of the incident Hg^{2+} ions when they enter the field-free cylinder C_2. To do this we used the modified form of the apparatus shown in fig. 1*b* and carried out the procedure described on p. 439. This gives us the energy distribution of the positive ions formed between G_1 and G_2, 90 % of which are Hg^+ ions and only 10 % Hg^{2+} ions for $V_0 = 60$ volts (Bleakney 1930). We cannot separate the Hg^{2+} ions from the Hg^+ ions, but, since they form only 10 % of the total, the curves obtained will represent the energy distribution of the Hg^+ ions. These curves, for the four different values of the accelerating potential V_1 used for the results given in fig. 4, are shown in fig. 5. The factors introducing the inhomogeneity into the Hg^+ ions will operate also on the Hg^{2+} ions, and thus we may take these curves to represent the energy distribution of the Hg^{2+} ions provided we multiply the abscissa scale of energy by the factor 2, since the Hg^{2+} ions have twice the energy of the Hg^+ ions owing to their double charge. Thus the energy spread of the Hg^{2+} ions is twice that of the Hg^+ ions. The lower energy scale in fig. 5 is for the Hg^+ ions which do not further concern us, while the upper scale is for the Hg^{2+} ions.

Points taken from these energy distribution curves for the incident Hg^{2+} ions have been reduced to fit the peaks of fig. 4 at the maxima and are plotted on these peaks. We see that for each value of V_1 the energy spread of the peaks in fig. 4 is fully accounted for by the energy spread in the incident Hg^{2+} ions. As shown above, the meaning of this is that the electron transfer process

$$Hg^{2+} + Hg \rightarrow Hg^+ + Hg^+$$

is far more probable, or perhaps only possible, for a collision in which one of the Hg^+ ions carries off all the kinetic energy of the incident Hg^{2+} ion, that is, for either a head-on or a grazing collision.

Probability of the electron transfer process

For any one value of V_1 let A be the area of the peak of emergent ions shown in fig. 4, and let a be the area of the peak of incident Hg^{2+} ions. From Bleakney's (1930) results the ratio of the number of Hg^{2+} ions formed by electrons of 60 electron-volts energy to the number of Hg^+ ions is 0·10, and therefore a is one-tenth the area of the peaks in fig. 5 which were obtained for the incident Hg^+ ions. The scale factor for the peaks of fig. 5 is the same as that for the peaks of fig. 4. The areas A and a are those of the hatched portion of the peaks in figs. 4 and 5. These cover an energy spread of 30 electron-volts in both figures. It should be remembered that the upper energy scale is the appropriate one in fig. 5 for the Hg^{2+} ions.

The fraction of the total number of incident Hg^{2+} ions which, on collision with neutral atoms, give $Hg^+ + Hg^+$ by electron transfer in passing through the field-free cylinder C_2 is then given by

$$N = (A/a)\,SIF, \tag{1}$$

where $S = 1/100$ is the ratio of the galvanometer shunt factors for the results of figs. 4 and 5 respectively, and $I = 1/8$ is the ratio of the filament emissions for the results of figs. 5 and 4 respectively. The latter was 1 mA for the results of fig. 5, and 8 mA for those of fig. 4. The ion current was found to be a linear function of the filament emission. The factor $F = 1/0{\cdot}71$ is introduced, since the number of gauzes in the apparatus of fig. 1*a* which gives the results of fig. 4 is one more than in the apparatus of fig. 1*b* which gives the results of fig. 5, the transmission area of each gauze being 71 % of its total area.

The cross-section for the electron transfer process is then

$$Q = N/nxp, \tag{2}$$

where $n = 3{\cdot}56 \times 10^{16}$ is the number of atoms per c.c. at a pressure of 1 mm. of Hg at 0° C, $p = 10^{-3}$ is the mercury vapour pressure in C_2 in mm. of Hg at 0° C, $x = 1{\cdot}5$ cm. is the distance the Hg^{2+} ions travel in the field-free cylinder C_2, and N is given by (1). From (1) and (2) we obtain

$$Q = 3{\cdot}3 \times 10^{-17} A/a.$$

The cross-section for the electron transfer process $Hg^{2+} + Hg \rightarrow Hg^{+} + Hg^{+}$ as a function of the energy of the incident Hg^{2+} ions in electron-volts is shown in fig. 6. The kinetic theory cross-section for collision of two neutral Hg atoms determined from viscosity measurements given in the Landolt-Börnstein Tables is $57{\cdot}7 \times 10^{-16}$ cm.^{2}. Therefore the cross-section for the electron transfer process over the energy range shown in fig. 6 is of the order of one-tenth the gas-kinetic cross-section for collision.

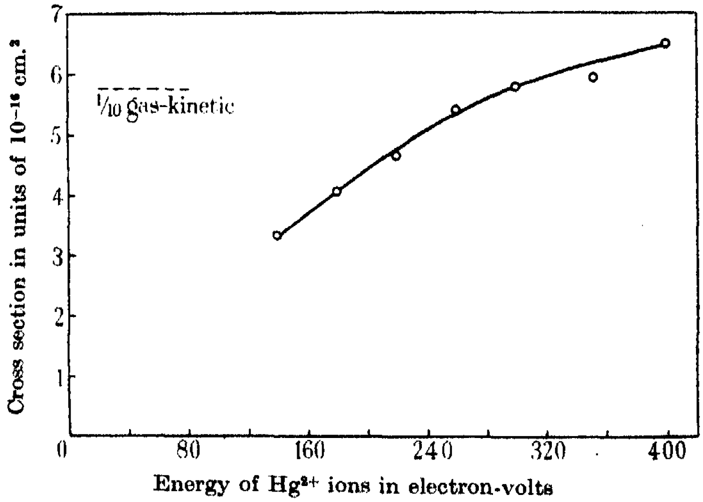


Fig. 6. Cross-section for electron transfer as a function of the energy of the incident Hg^{2+} ions. Broken line at 5·77 cm.^{2} is one-tenth the gas-kinetic cross-section for collision.

The mechanics of the collision process

We have seen from the results given in fig. 4 that the ions resulting from electron transfer all occur at a value of the retarding potential almost precisely equal to $2V_1$, and that the width of this peak is entirely accounted for by the energy spread in the incident Hg^{2+} ions. These results require the effective collision, on the basis of classical mechanics, to be either a head-on one or one in which the Hg^{2+} ion just grazes the neutral atom. We shall now investigate the distribution of the internal energy set free in the process amongst the two resulting Hg^{+} ions for these two types of collision.

(1) *Head-on collision.*

In the process $Hg^{2+} + Hg \rightarrow Hg^{+} + Hg^{+}$, let the velocity of the incident Hg^{2+} ion be v_0, and that of the two Hg^{+} ions resulting from the collision be v_1 and v_2. The velocity of the neutral Hg atom will be taken as zero, since it has only thermal energy. The mass m of all the particles concerned is the same. Let p be the internal energy set free by the electron transfer, which we shall first assume is all shared in the form of kinetic energy between the two resulting Hg^{+} ions. For a head-on collision we then have

$$\tfrac{1}{2}mv_0^2 + p = \tfrac{1}{2}mv_1^2 + \tfrac{1}{2}mv_2^2, \tag{3}$$

$$mv_0 = mv_1 + mv_2. \tag{4}$$

Squaring (4) we then have from (3)

$$v_1 v_2 = -p/m. \tag{5}$$

From (3) we then obtain

$$\tfrac{1}{2}mv_1^2 = \tfrac{1}{2}mv_0^2 - p^2/2mv_1^2 + p. \tag{6}$$

Writing $E = \tfrac{1}{2}mv_1^2$ and $E_0 = \tfrac{1}{2}mv_0^2$ we have from (6)

$$E^2 - (E_0 + p)\,E + p^2/4 = 0,$$

and hence $$E = \tfrac{1}{2}E_0 + \tfrac{1}{2}p \pm \tfrac{1}{2}E_0(1 + 2p/E_0)^{\frac{1}{2}}.$$

Since $p = 8{\cdot}28$ electron-volts (see p. 441) and E_0 is always greater than 140 electron-volts we have

$$E = \tfrac{1}{2}E_0 + \tfrac{1}{2}p \pm \tfrac{1}{2}E_0(1 + p/E_0).$$

The two solutions for E are therefore

$$\left.\begin{aligned} E &= E_0 + p, \\ E &= 0. \end{aligned}\right\} \tag{7}$$

Thus one of the Hg^{+} ions travels forward with kinetic energy equal to that of the incident Hg^{2+} ion plus the total internal energy set free in the collision, while the other Hg^{+} ion remains at rest. The energy $E_0 = 2V_1$.

(2) *Grazing collision.*

When the Hg^{2+} ion just grazes the Hg atom the internal energy set free by the electron transfer will affect the relative motion of the particles only in a direction at right angles to the direction of motion of the Hg^{2+} ion.

Let v_1 and v_2 now be the velocities of the two resulting Hg^+ ions in a direction at right angles to the direction of motion of the Hg^{2+} ion. We now have

$$p = \tfrac{1}{2}mv_1^2 + \tfrac{1}{2}mv_2^2, \tag{8}$$

$$0 = mv_1 + mv_2. \tag{9}$$

As before, we obtain $$v_1 v_2 = -p/m,$$

and hence from (8) $$p = \tfrac{1}{2}mv_1^2 + p^2/2mv_1^2. \tag{10}$$

Writing $E = \frac{1}{2}mv_1^2$ we have from (10)

$$E^2 - pE + p^2/4 = 0,$$

and hence $$E = \tfrac{1}{2}p. \tag{11}$$

Thus the internal energy set free in the process is shared equally between the two Hg^+ ions, causing motion in a direction at right angles to the direction of motion of the incident Hg^{2+} ion. The internal energy set free does not affect the component velocity of the resulting Hg^+ ions in the direction of motion of the incident Hg^{2+} ion.

From (7) and (11) we see that the fast Hg^+ ion resulting from the electron transfer process should have a component velocity in the direction of the retarding field corresponding to an energy of $2V_1 + p$ if the collision is head-on, and of only $2V_1$ if the collision is a grazing one. We should therefore be able to decide from the position of the maximum of the peaks in fig. 4 whether the effective collision is a head-on one or a grazing one.

The first column of Table I contains the seven values of V_1 used. The second column gives the energy of the fast Hg^+ ions resulting from the electron transfer process, determined from the position of the maximum of the peaks, four of which are shown in fig. 4. The third column gives the energy of these ions in excess of $2V_1$.

TABLE I

V_1 volts	Energy of Hg^+ ions electron-volts	Energy in excess of $2V_1$ electron-volts
70	140	0
90	183	3·0
110	222	2·0
130	262·5	2·5
150	301	1·0
176	353·5	1·5
200	402	2·0

It appears from this table that the fast Hg^+ ions have an energy in excess of $2V_1$ by about 2 electron-volts, but definitely not as great as $p = 8{\cdot}28$ electron-volts. Therefore the collision cannot be a grazing one, nor can it be a head-on one in which the whole of the internal energy is converted into kinetic energy. We conclude therefore that the effective collision resulting in the electron transfer process is a head-on collision in which part of the 8·28 electron-volts internal energy is employed in exciting one of the resulting Hg^+ ions, leaving only about 2 electron-volts to be added as kinetic energy to the fast Hg^+ ion. The energy available for excitation is thus $8{\cdot}28 - 2 = 6{\cdot}3$ electron-volts. There are two excited states of the Hg^+ ion having an energy of about this amount, the $5d^9\,6s^2$, $^2\mathrm{D}_{1\frac{1}{2}}$ state of 6·23 electron-volts energy and the $5d^{10}\,6p$, $^2\mathrm{P}_{\frac{1}{2}}$ state of 6·35 electron-volts energy.

One would expect the greater part of the 8·28 electron-volts internal energy to be employed in excitation on grounds other than the experimental ones given above, for theoretically an electron transfer process is only probable for close resonance. In general, the smaller the amount of internal energy to be converted into kinetic energy the greater the probability of electron transfer taking place during the collision.

The process shown to occur in this work can now be visualized as follows. An Hg^{2+} ion collides head-on with a stationary Hg atom. An electron from the Hg atom is captured by the Hg^{2+} ion into an excited state of about 6·3 electron-volts energy, leaving both particles singly charged. The Hg^+ ion formed from the Hg^{2+} ion remains stationary, and the Hg^+ ion formed from the neutral atom moves forward in the direction of the incident Hg^{2+} ion. Resulting from the electron transfer an amount of energy equal to the difference between the first and second ionization potentials of the Hg atom less the energy of the excited state of the Hg^+ ion is set free and must be shared as kinetic energy between the separating particles. The mechanics of the head-on collision shows that this internal energy, which is about 2 electron-volts, is absorbed entirely by the fast Hg^+ ion, provided it is small compared to the kinetic energy of the incident Hg^{2+} ion. Thus the energy of the fast Hg^+ ion resulting from the collision is equal to that of the incident Hg^{2+} ion plus about 2 electron-volts.

We wish to thank our colleague, Dr R. A. Smith, for many enlightening discussions during the course of this work.

SUMMARY

An investigation of the electron transfer process

$$Hg^{2+} + Hg \rightarrow Hg^{+} + Hg^{+}$$

has been made using incident Hg^{2+} ions having energies from 140 to 400 electron-volts. The internal energy set free in the collision by the capture of the electron is 8·[illegible] electron-volts, of which it is found that about 6·3 electron-volts are employed in exciting one of the resulting Hg^{+} ions, the balance of about 2 electron-volts being shared as kinetic energy between the two Hg^{+} ions in accordance with the law of conservation of momentum.

The classical mechanics of the collision process show that the results obtained can be accounted for only by a head-on collision, one of the resulting Hg^{+} ions remaining stationary and the other carrying off the full kinetic energy of the incident Hg^{2+} ion plus the total amount of 2 electron-volts internal energy available in the form of kinetic energy.

The cross-section for the process [illegible] energy range of 140 to 400 electron-volts has been obtained, and t[illegible]crease of cross-section with energy [illegible] incident Hg^{2+} ion, the value [illegible]ddle of the energy range bein[illegible] order of one-tenth of the gas-[illegible] [illegible]tion for collision.

REFERENCES

Bacher and Goudsmit 1932 "Atomic Energy States", p. 231. McGraw-Hill Book Co.

Bleakney 1930 *Phys. Rev.* **35**, 1180.

Harnwell 1927 *Phys. Rev.* **29**, 683, 830.

Henderson 1922 *Proc. Roy. Soc* A, **102**, 496.

Hogness and Lunn 1927 *Phys. Rev.* **30**, 26.

Kallmann and Rosen 1930 *Z. Phys.* **64**, 806.

Rutherford, Chadwick and Ellis 1930 "Radiations from Radioactive Substances" p. 119. Camb. Univ. Press.

A new Coriolis perturbation in the methane spectrum
III. Intensities and optical spectrum

By W. H. J. Childs and H. A. Jahn
Davy-Faraday Laboratory, The Royal Institution

(*Communicated by Sir William Bragg, O.M., P.R.S.—Received* 22 *June* 1938—*Revised* 22 *November* 1938)

[Plate 28]

Introduction

In two previous papers (Jahn 1938) the perturbations of the rotational levels of the low-frequency fundamental vibration ν_4 of the methane molecule were evaluated. These perturbations were caused by the Coriolis term in the vibrational-rotational Hamiltonian which couples the rotational levels of this fundamental with the rotational levels of the infra-red inactive vibration ν_2. In the present paper we use this energy spectrum to calculate the theoretical infra-red absorption spectrum of the vibration ν_4. For this purpose we need the intensities of the rotational fine structure lines and these are here calculated, taking into account the nuclear spin weights of the states. We do not, however, take into account the effect of the Coriolis perturbation itself on the intensities, since the perturbation of the levels compared with the difference of their total vibrational energy is very small indeed, so that any intermixture of the wave functions of the two sets of levels can be neglected. From the theoretical spectrum so obtained we calculate the absorption envelope which would be observed with slit widths of approximately 0·5 cm.$^{-1}$ and compare this with the experimental envelope. Attention is drawn to the fact that a third overtone absorption band of methane has a complex rotational structure which is very similar indeed to that of the fundamental here under discussion, and the reason for this is given. The general importance of Coriolis perturbations in the spectra of polyatomic molecules is stressed, and the general nature of the problems which will be met with in the rotational analysis of the spectra of these molecules is outlined.

1. Intensities

Before we derive the theoretical intensities of the fine structure lines arising from transitions to the Coriolis sublevels, we consider the theoretical intensities in a simple absorption band of methane where there are no interaction effects. It would appear that the theoretical intensities in this simple case have not been given correctly before. These intensities depend upon three factors: first, upon the transition amplitude or sum of matrix elements squared of the electric moment from one single state; secondly, upon the statistical weight of the initial energy level, this weight being determined by the rotational and nuclear spin degeneracy; finally, upon the Boltzmann factor of this level. It is customary to introduce an intensity or i-factor defined as the product of the transition amplitude and the degree of degeneracy the initial level would have in the absence of nuclear spin. This i-factor must then be multiplied by a correcting nuclear spin factor g and by the Boltzmann factor to give the theoretical line strength.

The transition amplitudes from a single (J, K, M) state of the symmetrical top are well known and are given below (Table I). They are obtained by summation of the squares of the matrix elements of the direction cosines over the transitions $M \to M+1$, M, $M-1$ of the magnetic quantum number. From these amplitudes we obtain the transition amplitudes of

TABLE I

Transition		Amplitude	Total
$J \to J-1$	$K \to K-1$	$\frac{(J+K)(J-1+K)}{2J(2J+1)}$	$\frac{2J-1}{2J+1}$
	$K \to K$	$\frac{J^2-K^2}{J(2J+1)}$	
	$K \to K+1$	$\frac{(J-K)(J-1-K)}{2J(2J+1)}$	
$J \to J$	$K \to K-1$	$\frac{(J+K)(J+1-K)}{2J(J+1)}$	1
	$K \to K$	$\frac{K^2}{J(J+1)}$	
	$K \to K+1$	$\frac{(J-K)(J+1+K)}{2J(J+1)}$	
$J \to J+1$	$K \to K-1$	$\frac{(J+1-K)(J+2-K)}{2(J+1)(2J+1)}$	$\frac{2J+3}{2J+1}$
	$K \to K$	$\frac{(J+1)^2-K^2}{(J+1)(2J+1)}$	
	$K \to K+1$	$\frac{(J+1+K)(J+2+K)}{2(J+1)(2J+1)}$	

the spherical top by summation over the transitions in the quantum number K as shown in the table. Since the initial level is $(2J+1)^2$ fold degenerate we obtain the intensity factors for the spherical top shown in Table II.

TABLE II. INTENSITY FACTORS FOR SPHERICAL TOP

J	$R(J)$	$Q(J)$	$P(J)$
	$(2J+1)(2J+3)$	$(2J+1)^2$	$(2J+1)(2J-1)$
0	3	—	—
1	15	9	3
2	35	25	15
3	63	49	35
4	99	81	63
5	143	121	99
6	195	169	143
7	255	225	195
8	323	289	255
9	399	361	323
10	483	441	399
11	575	529	483
12	675	625	575
13	783	729	675
14	899	841	783
15	1023	961	899

The nuclear spin factors for CH_4 can be obtained in the following manner (cf. Bright Wilson 1935, A.-W. Maue 1937). The initial rotational levels can be classified according to the irreducible representations A, E, F of the group T of all those rotations about axes fixed in the molecule which permute the identical hydrogen nuclei. Those rotational states which transform according to the representation A have an extra nuclear spin weight of 5, whilst those which transform according to the representations E and F have no extra weights. Thus if the reduction of the rotational wave functions u_K^J of the Jth rotational level is given by

$$D_J = aA + bE + cF$$

with $a + 2b + 3c = 2J + 1$ (this reduction being independent of the magnetic quantum number M) then the nuclear spin factor of this level is

$$g_J = \frac{5a + 2b + 3c}{2J+1}.$$

The spin factors g obtained in this way are shown in Table III, together with the Boltzmann factors calculated for $kT = 200$ cm.$^{-1}$ ($t = 16°$ C) and with $B'' = 5{\cdot}27$ cm.$^{-1}$.

TABLE III. SPIN FACTORS AND BOLTZMANN FACTORS FOR CH_4

($B'' = 5{\cdot}27$ cm.$^{-1}$, $kT = 200$ cm.$^{-1}$)

J	Term system	Spin factor g	E	$e^{-E/kT}$	$ge^{-E/kT}$
0	A	5	0	1	5
1	F	1	10·54	0·9484	0·9484
2	$E+F$	1	31·62	0·8538	0·8538
3	$A+2F$	$\frac{11}{7} = 1{\cdot}571$	63·24	0·7290	1·145
4	$A+E+2F$	$\frac{13}{9} = 1{\cdot}444$	105·40	0·5904	0·8525
5	$E+3F$	1	158·10	0·4533	0·4533
6	$2A+E+3F$	$\frac{21}{13} = 1{\cdot}615$	221·34	0·3306	0·5339
7	$A+E+4F$	$\frac{19}{15} = 1{\cdot}267$	295·12	0·2285	0·2895
8	$A+2E+4F$	$\frac{21}{17} = 1{\cdot}235$	379·44	0·1500	0·1853
9	$2A+E+5F$	$\frac{27}{19} = 1{\cdot}421$	474·30	0·09329	0·1326
10	$2A+2E+5F$	$\frac{29}{21} = 1{\cdot}381$	579·70	0·05508	0·07607
11	$A+2E+6F$	$\frac{27}{23} = 1{\cdot}174$	695·64	0·03087	0·03624
12	$3A+2E+6F$	$\frac{37}{25} = 1{\cdot}480$	822·12	0·01638	0·02424
13	$2A+2E+7F$	$\frac{35}{27} = 1{\cdot}296$	959·14	0·008263	0·01071
14	$2A+3E+7F$	$\frac{37}{29} = 1{\cdot}276$	1106·70	0·003952	0·005043
15	$3A+2E+8F$	$\frac{43}{31} = 1{\cdot}387$	1264·80	0·001793	0·002487

From Tables II and III we calculate the theoretical relative line strengths shown in Table IV. These relative intensities are shown in fig. 1 for the *P* and *R* branches.

TABLE IV. THEORETICAL LINE STRENGTHS FOR CH_4

J	$R(J)$	$Q(J)$	$P(J)$
0	15·00	—	—
1	14·23	8·536	2·845
2	29·88	21·35	12·81
3	72·14	56·11	40·08
4	84·40	69·05	53·71
5	64·82	54·85	44·88
6	104·1	90·23	76·35
7	73·82	65·14	56·45
8	59·85	53·55	47·25
9	52·91	47·87	42·83
10	36·74	33·55	30·35
11	20·84	19·17	17·50
12	16·36	15·15	13·94
13	8·386	7·808	7·229
14	4·534	4·241	3·949
15	2·544	2·390	2·236

When the upper state rotational levels are split by the Coriolis perturbation the total line strength is distributed over the fine structure lines in a manner determined by the spin weights of the individual initial symmetry sublevels from which they originate. The theoretical intensities of the fine structure lines are thus easy to calculate taking into account the additional spin weight of 5 for the sublevels of type A. (The representations A_1, A_2 of the full tetrahedral group T_d give the same representation A of the tetrahedral group T of pure rotations, similarly F_1 and F_2 of T_d are to be identified with F of T.) When two or more sublevels of the same symmetry type occur for a given value of J, the corresponding intensity is distributed equally

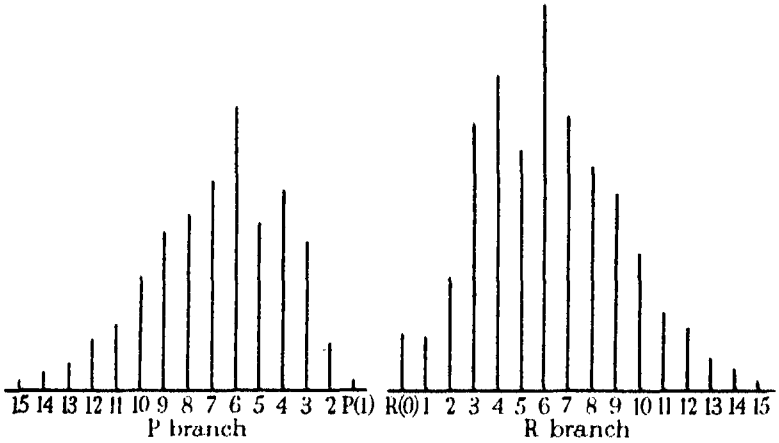


FIG. 1. Theoretical line strengths for CH_4.

TABLE V. THEORETICAL LINE STRENGTHS FOR THE FINE STRUCTURE LINES

J	$R(J)$			$Q(J)$			$P(J)$		
	A_1 or A_2	E	F_1 or F_2	A_1 or A_2	E	F_1 or F_2	A_1 or A_2	E	F_1 or F_2
0	15	—	—	—	—	—	—	—	—
1	—	—	14·23	—	—	8·54	—	—	2·85
2	—	11·95	17·93	—	8·54	12·81	—	5·12	7·69
3	32·82	—	19·69	25·52	—	15·31	18·23	—	10·94
4	32·47	12·99	19·48	26·57	10·63	15·94	20·67	8·27	12·40
5	—	11·79	17·68	—	9·97	14·96	—	8·16	12·24
6	24·80	9·92	14·88	21·49	8·60	12·89	18·19	7·27	10·91
7	19·42	7·77	11·65	17·14	6·86	10·28	14·86	5·94	8·91
8	14·25	5·70	8·55	12·75	5·10	7·65	11·25	4·50	6·75
9	9·80	3·92	5·88	8·87	3·55	5·32	7·93	3·17	4·76
10	6·34	2·53	3·80	5·79	2·31	3·47	5·24	2·09	3·14

over these. The theoretical intensities for the fine structure lines calculated in this manner are shown in Table V, where the division necessitated by the occurrence of lines of the same symmetry type is already carried out.

2. THEORETICAL SPECTRUM

To calculate the optical spectrum we need the energies of the rotational levels of ν_4. Starting from the energies of Teller's unperturbed Coriolis levels (taking $B = 5{\cdot}27$ cm.$^{-1}$, $\zeta_4 = 0{\cdot}45$) and subtracting the perturbations evaluated in Part II, we calculated the energies of the Coriolis sublevels.

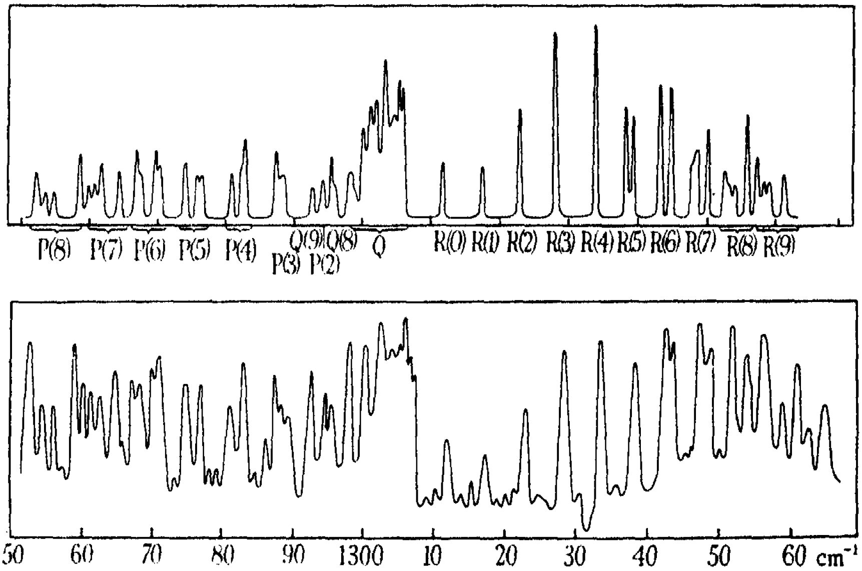


FIG. 2. Below: Envelope of the 1306 cm.$^{-1}$ fundamental vibrational band of methane as observed by Nielsen and Nielsen. Above: Theoretical envelope showing fine structure arising from the Coriolis perturbation.

From these energy levels we found the line frequencies given in Table VI, where we have taken $\nu_0 = 1306{\cdot}00$ cm.$^{-1}$ and have assumed the same value of $B(5{\cdot}27$ cm.$^{-1})$ for the upper and ground states. The envelope of these fine structure lines obtained by replacing each line by a triangle with base of 1 cm.$^{-1}$ (corresponding to slit widths of $0{\cdot}5$ cm.$^{-1}$) and height proportional to the intensity is shown in fig. 2, the areas of overlapping triangles being added together. In the same figure we show the experimental envelope of Nielsen and Nielsen (1935). The agreement between the two is remarkably good, particularly in the P branch. Exact agreement can only be expected

TABLE VI. FREQUENCIES OF FINE STRUCTURE LINES

Line	Intensity	Frequency	Line	Intensity	Frequency	Line	Intensity	Frequency
$R(0)\ A_1$	15	1311·80	$Q(1)\ F_1$	9	1305·90	$P(1)\ F_2$	3	1300·20
$R(1)\ F_2$	14	1317·42	$Q(2)\ F_1$	13	1305·70	$P(2)\ E$	5	1294·41
			$Q(2)\ E$	9	1305·40	$P(2)\ F_2$	8	1294·10
$R(2)\ E$	12	1322·99						
$R(2)\ F_2$	18	1322·90	$Q(3)\ A_2$	26	1306·00	$P(3)\ F_2$	11	1288·48
			$Q(3)\ F_2$	15	1305·25	$P(3)\ F_1$	11	1287·97
$R(3)\ F_2$	20	1328·50	$Q(3)\ F_1$	15	1304·65	$P(3)\ A_1$	18	1287·33
$R(3)\ F_1$	20	1328·31						
$R(3)\ A_1$	33	1328·08	$Q(4)\ F_1$	16	1305·63	$P(4)\ A_1$	21	1282·81
			$Q(4)\ E$	11	1304·28	$P(4)\ F_1$	12	1282·32
$R(4)\ A_1$	32	1334·09	$Q(4)\ F_2$	16	1303·83	$P(4)\ E$	8	1281·97
$R(4)\ F_1$	19	1333·88	$Q(4)\ A_2$	27	1303·20	$P(4)\ F_2$	12	1280·91
$R(4)\ E$	13	1333·74						
$R(4)\ F_2$	19	1333·29	$Q(5)\ F_1^{(1)}$	15	1305·31	$P(5)\ F_2^{(1)}$	12	1276·64
			$Q(5)\ E$	10	1304·92	$P(5)\ F_1$	12	1276·04
$R(5)\ F_2^{(1)}$	18	1339·46	$Q(5)\ F_2$	15	1302·82	$P(5)\ E$	8	1274·48
$R(5)\ F_1$	18	1339·20	$Q(5)\ F_1^{(2)}$	15	1302·01	$P(5)\ F_2^{(2)}$	12	1274·19
$R(5)\ E$	12	1338·48						
$R(5)\ F_2^{(2)}$	18	1338·36	$Q(6)\ A_2$	21	1305·40	$P(6)\ E$	7	1270·62
			$Q(6)\ F_2$	13	1304·76	$P(6)\ F_2^{(1)}$	11	1270·44
$R(6)\ E$	10	1344·92	$Q(6)\ F_1^{(1)}$	13	1304·03	$P(6)\ A_2$	18	1269·88
$R(6)\ F_2^{(1)}$	15	1344·83	$Q(6)\ A_1$	9	1301·29	$P(6)\ F_2^{(2)}$	11	1267·88
$R(6)\ A_2$	25	1344·56	$Q(6)\ F_1^{(2)}$	13	1300·51	$P(6)\ F_1$	11	1267·36
$R(6)\ F_2^{(2)}$	15	1343·60	$Q(6)\ E$	9	1300·26	$P(6)\ A_1$	18	1266·89
$R(6)\ F_1$	15	1343·35						
$R(6)\ A_1$	25	1343·12	$Q(7)\ F_1^{(1)}$	10	1304·88	$P(7)\ F_2^{(1)}$	9	1264·56
			$Q(7)\ E$	7	1303·82	$P(7)\ F_1^{(1)}$	9	1264·24
$R(7)\ F_2^{(1)}$	12	1350·36	$Q(7)\ F_2^{(1)}$	10	1303·17	$P(7)\ A_1$	15	1261·81
$R(7)\ F_1^{(1)}$	12	1350·21	$Q(7)\ A_2$	17	1301·91	$P(7)\ F_1^{(2)}$	9	1260·91
$R(7)\ A_1$	19	1349·02	$Q(7)\ F_2^{(2)}$	10	1298·54	$P(7)\ E$	6	1260·43
$R(7)\ F_1^{(2)}$	12	1348·58	$Q(7)\ F_1^{(2)}$	10	1298·11	$P(7)\ F_2^{(2)}$	9	1259·68
$R(7)\ E$	8	1348·34						
$R(7)\ F_2^{(2)}$	12	1347·97	$Q(8)\ F_1^{(1)}$	8	1304·48	$P(8)\ A_1$	11	1258·61
			$Q(8)\ E^{(1)}$	5	1304·01	$P(8)\ F_1^{(1)}$	7	1258·40
$R(8)\ A_1$	14	1355·86	$Q(8)\ F_2^{(1)}$	8	1302·44	$P(8)\ E^{(1)}$	5	1258·27
$R(8)\ F_1^{(1)}$	9	1355·75	$Q(8)\ F_1^{(2)}$	8	1300·67	$P(8)\ F_2^{(1)}$	7	1254·83
$R(8)\ E^{(1)}$	6	1355·69	$Q(8)\ E^{(2)}$	5	1295·89	$P(8)\ F_1^{(2)}$	7	1253·55
$R(8)\ F_2^{(1)}$	9	1353·99	$Q(8)\ F_2^{(2)}$	8	1295·71	$P(8)\ E^{(2)}$	5	1252·42
$R(8)\ F_1^{(2)}$	9	1353·36	$Q(8)\ A_2$	13	1295·40	$P(8)\ F_2^{(2)}$	7	1252·09
$R(8)\ E$	6	1352·81						
$R(8)\ F_2^{(1)}$	9	1352·64	$Q(9)\ A_2$	9	1304·44	$P(9)\ F_2^{(1)}$	5	1252·47
			$Q(9)\ F_2^{(1)}$	5	1303·84	$P(9)\ F_1^{(1)}$	5	1251·99
$R(9)\ F_2^{(1)}$	6	1361·27	$Q(9)\ F_1^{(1)}$	5	1303·08			
$R(9)\ F_1^{(1)}$	6	1361·18	$Q(9)\ A_1$	9	1301·24			
$R(9)\ E$	4	1359·05	$Q(9)\ F_1^{(2)}$	5	1299·31			
$R(9)\ F_2^{(2)}$	6	1358·91	$Q(9)\ E$	4	1298·84			
$R(9)\ A_2$	10	1358·20	$Q(9)\ F_2^{(2)}$	5	1292·84			
$R(9)\ F_2^{(3)}$	6	1357·50	$Q(9)\ F_1^{(3)}$	5	1292·59			
$R(9)\ F_1^{(2)}$	6	1357·27						
$R(9)\ A_1$	10	1357·06						

when vibrational and centrifugal changes in the equilibrium configuration and other perturbations are taken into account. This will be particularly the case for the envelope of the Q branch which is very sensitive to slight changes in the levels. It should be pointed out that the Coriolis coupling between the rotational levels of the low frequency fundamental ν_4 and those of the other active fundamental ν_3 may produce perturbations of the order of 1 cm.$^{-1}$ for $J = 10$, in spite of the fact that the levels concerned are separated by over 1700 cm.$^{-1}$. Thus an exact treatment would have to take this effect into account.

In fig. 3 (Plate 28) we show also a reproduction of the R branch of the 9047 cm.$^{-1}$ overtone band of methane as observed in the photographic infra-red (Childs 1936), together with its corresponding photometer trace. The reproduction shows the intensity changes very clearly. It will be seen that there is a remarkable similarity between the rotational fine structure of this overtone and that of the fundamental here under discussion. This can be understood if it is realized that not only is the methane molecule a regular tetrahedron in the ground state, but that its vibrational-rotational energy levels cannot lose their regular tetrahedral symmetry no matter how large or anharmonic the vibrations become. This is a consequence of the fact that the total Hamiltonian of the molecule has tetrahedral symmetry, so that therefore the anharmonic terms, the centrifugal terms and any Coriolis terms can all be regarded as tetrahedrally symmetrical perturbations. They cannot therefore remove any further the degeneracy of the sublevels considered above. Thus the same type and number of levels will be involved in any active overtone of methane. The appearance of the band will however depend upon the magnitude of the resultant perturbation to which the levels are subject. In the simple overtone band we reproduce, the perturbations must to a large extent cancel out leaving a small resultant perturbation similar to that which the levels of the fundamental band undergo. If the perturbation is very large, however, the fine structure lines can be spread out over the whole band and an apparently complicated structure will result. When one realizes that the Coriolis coupling terms can give rise to large perturbations,* it is not to be wondered at that some of the methane overtone bands have a very complicated structure. It should be emphasized however that the levels always remain those of a regular tetrahedral spherical top and consequently there will be no occasion to consider any further splitting of the levels involved (so long as there is no

---

* The perturbation of the type considered here clearly increases with the vibrational quantum number. Calculations of the rotational structure of overtones and combination bands are in progress and will be reported on later.

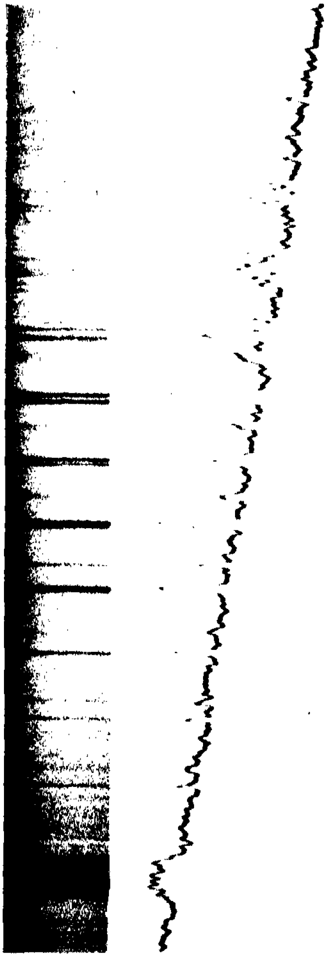


FIG. 3. Above: reproduction of the *R* branch of the 9047 cm.$^{-1}$ overtone band of methane as observed in the photographic infra-red. The reproduction is from a positive process copy of the original plate, hence the absorption lines appear as dark lines on a clear background. Below: photometer trace of the same process plate.

perturbation external to the molecule). Thus the bands shown here can be regarded as typical spectra of a regular tetrahedral molecule. Applications to the spectra of the germane and silane molecules have been mentioned elsewhere (Jahn and Childs 1938).

CONCLUSION

It is clear from this investigation that Coriolis perturbations between different bands of polyatomic molecules can have a very important effect on the observed infra-red (and Raman) spectrum of the molecule. Such perturbations are possible in all molecules which have a threefold or higher axis of symmetry. It is a general result also that the nature of the splitting of the vibrational-rotational levels of a molecule is determined by the symmetry of the molecular equilibrium configuration. Thus the analysis of the spectrum of any polyatomic molecule will involve considerations of the same nature as have been here developed for the special case of the methane molecule.

The authors are indebted to the Director and the Managers of the Royal Institution for the facilities afforded at the Davy-Faraday Laboratory where this research was carried out.

SUMMARY

This paper completes an explanation of the complicated fine structure of the infra-red absorption band of methane at 1306 cm.$^{-1}$, in terms of a tetrahedral configuration of the molecule. In Parts I and II it is shown that the rotational levels of this low frequency fundamental ν_4 are subject to perturbations from the inactive fundamental ν_2 at 1536 cm.$^{-1}$. These arise because the rotational levels of these two vibrations are coupled by a Coriolis term in the Hamiltonian. These perturbations were evaluated up to the tenth rotational quantum number, and in conjunction with calculations of the intensities of the rotational fine structure lines we derive in the present paper the theoretical envelope of the absorption band. This theoretical spectrum agrees very well with the observations of Nielsen and Nielsen (see fig. 2).

An overtone band of methane found at 9047 cm.$^{-1}$ in the photographic infra-red shows a similar complex structure (fig. 3). This is readily understood when one realizes that the methane molecule cannot lose its tetrahedral symmetry no matter how large or anharmonic the vibrations become, so

that the same types and number of rotational sublevels will be involved in any active overtone of methane.

This new effect may be expected in the spectra of all polyatomic molecules possessing a threefold or higher axis of symmetry and in the Appendix we give a general method of calculating for any given molecule which perturbing terms of this kind can occur in the Hamiltonian.

APPENDIX

Coriolis terms in polyatomic molecules

In this and two previous papers we have shown that there is a Coriolis term in the Hamiltonian of the methane molecule which couples the rotational levels of two different normal modes of vibration and has a very important effect on the infra-red absorption spectrum of the molecule. The question arises whether one can predict what Coriolis terms are possible in the vibrational-rotational Hamiltonian of any given molecule. This general problem has indeed been propounded by Eckart (1935). It can be solved immediately by the help of group theory. The Coriolis terms involve vibrational angular momenta, which are obtained by combining the vibrational normal co-ordinates with the corresponding impulses. Now these normal co-ordinates are transformed by the elements of the symmetry group of the molecule according to definite irreducible representations of the group, in a way which has been investigated by Tisza (1933). The vibrational impulses are transformed in the same way as the corresponding co-ordinates. On the other hand, the transformation properties of the components of vibrational angular momentum are determined completely by the fact that they are the components of an antisymmetrical tensor (or axial vector). Thus knowing the representations of all the normal co-ordinates of the molecule, the problem reduces to that of finding which products of these representations contain the representation of the antisymmetrical tensor. Further the explicit reduction of the product representations in question will give us the explicit form of the vibrational angular momenta which can occur (and in general will occur) in the Hamiltonian of the molecule.

In the case of the methane molecule the nine normal co-ordinates transform in the following way according to irreducible representations of the symmetry group (T_d) of the molecule:

$$q_{1a} \sim A_1; \qquad \left.\begin{matrix} q_{2e} \\ q_{2f} \end{matrix}\right\} \sim E; \qquad \left.\begin{matrix} q_{3x} \\ q_{3y} \\ q_{3z} \end{matrix}\right\} \sim F_2; \qquad \left.\begin{matrix} q_{4x} \\ q_{4y} \\ q_{4z} \end{matrix}\right\} \sim F_2.$$

On the other hand, the representation of an antisymmetrical tensor is the antisymmetrical product representation $\{V^2\}$ of the representation V of a vector. For the tetrahedral group T_d we have

$$V = F_2,$$
$$\{V^2\} = \{F_2^2\} = F_1.$$

Thus the components of angular momentum in the case of methane transform according to the representation F_1. The following product representations of the normal co-ordinates and their impulses contain the representation F_1:

$$E \times F_2 = F_1 + F_2,$$
$$F_2 \times F_2 = A_1 + E + F_1 + F_2.$$

Denoting quantities which transform according to the representations E, F_1, F_2 by e, f; X, Y, Z; x, y, z respectively, one finds easily that the components which transform according to F_1 for the above product representations are as follows:

$$E \times F_2 \quad \begin{cases} (X) = \left(\tfrac{1}{2}e - \frac{\sqrt{3}}{2}f\right)x, \\ (Y) = \left(\tfrac{1}{2}e + \frac{\sqrt{3}}{2}f\right)y, \\ (Z) = -ez. \end{cases}$$

$$F_2^{(1)} \times F_2^{(2)} \quad \begin{cases} (X) = \frac{1}{\sqrt{2}}(y^{(1)}z^{(2)} - z^{(1)}y^{(2)}), \\ (Y) = \frac{1}{\sqrt{2}}(z^{(1)}x^{(2)} - x^{(1)}z^{(2)}), \\ (Z) = \frac{1}{\sqrt{2}}(x^{(1)}y^{(2)} - y^{(1)}x^{(2)}). \end{cases}$$

From these explicit reductions one finds the following possible vibrational angular momenta for the methane molecule:

$$S^{(23)} \quad \begin{cases} S_X^{(23)} = \zeta_{23}\left\{\left(\tfrac{1}{2}q_{2e} - \frac{\sqrt{3}}{2}q_{2f}\right)p_{3x} - \left(\tfrac{1}{2}p_{2e} - \frac{\sqrt{3}}{2}p_{2f}\right)q_{3x}\right\}, \\ S_Y^{(23)} = \zeta_{23}\left\{\left(\tfrac{1}{2}q_{2e} + \frac{\sqrt{3}}{2}q_{2f}\right)p_{3y} - \left(\tfrac{1}{2}p_{2e} + \frac{\sqrt{3}}{2}p_{2f}\right)q_{3y}\right\}, \\ S_Z^{(23)} = \zeta_{23}\{-q_{2e}p_{3z} + p_{2e}q_{3z}\}, \end{cases}$$

$$S^{(24)}\begin{cases} S^{(24)}_X = \zeta_{24}\left\{\left(\tfrac{1}{2}q_{2e}-\dfrac{\sqrt{3}}{2}q_{2f}\right)p_{4x}-\left(\tfrac{1}{2}p_{2e}-\dfrac{\sqrt{3}}{2}p_{2f}\right)q_{4x}\right\}, \\ S^{(24)}_Y = \zeta_{24}\left\{\left(\tfrac{1}{2}q_{2e}+\dfrac{\sqrt{3}}{2}q_{2f}\right)p_{4y}-\left(\tfrac{1}{2}p_{2e}+\dfrac{\sqrt{3}}{2}p_{2f}\right)q_{4y}\right\}, \\ S^{(24)}_Z = \zeta_{24}\{-q_{2e}p_{4z}+p_{2e}q_{4z}\}, \end{cases}$$

$$S^{(34)}\begin{cases} S^{(34)}_X = \zeta_{34}\{(q_{3y}p_{4z}-q_{3z}p_{4y})-(p_{3y}q_{4z}-p_{3z}q_{4y})\}, \\ S^{(34)}_Y = \zeta_{34}\{(q_{3z}p_{4x}-q_{3x}p_{4z})-(p_{3z}q_{4x}-p_{3x}q_{4z})\}, \\ S^{(34)}_Z = \zeta_{34}\{(q_{3x}p_{4y}-q_{3y}p_{4x})-(p_{3x}q_{4y}-p_{3y}q_{4x})\}, \end{cases}$$

$$S^{(3)}\begin{cases} S^{(3)}_X = \zeta_3(q_{3y}p_{3z}-q_{3z}p_{3y}), \\ S^{(3)}_Y = \zeta_3(q_{3z}p_{3x}-q_{3x}p_{3z}), \\ S^{(3)}_Z = \zeta_3(q_{3x}p_{3y}-q_{3y}p_{3x}), \end{cases}$$

$$S^{(4)}\begin{cases} S^{(4)}_X = \zeta_4(q_{4y}p_{4z}-q_{4z}p_{4y}), \\ S^{(4)}_Y = \zeta_4(q_{4z}p_{4x}-q_{4x}p_{4z}), \\ S^{(4)}_Z = \zeta_4(q_{4x}p_{4y}-q_{4y}p_{4x}), \end{cases}$$

where the p's are the impulses corresponding to the co-ordinates q, and the ζ's are numerical factors. $S^{(3)}$ and $S^{(4)}$ are the well-known vibrational angular momenta of the threefold degenerate vibrations considered by Teller. $S^{(24)}$ is the vibrational angular momentum which has been considered in this paper and which has an important effect on the spectrum because of the fact that the frequencies ν_2 and ν_4 of methane lie relatively close together. It can be verified easily that the remaining angular momenta $S^{(23)}$ and $S^{(34)}$ also exist in methane, assuming valency and deformation modes of vibration. (The numerical coefficients ζ depend upon the exact form of the vibrations and hence upon the force system. For special force systems they may vanish.) For valency and deformation modes we find

$$\zeta_{23} = \sqrt{\frac{2m_0}{2m_0+M}}, \quad \zeta_{34} = \frac{3}{\sqrt{2}}\frac{\sqrt{m_0 M}}{2m_0+M},$$

where m_0 = mass of carbon atom, M = mass of CH_4. The values of ζ_3, ζ_4 and ζ_{24} have been given in our papers.

Since the product of E with itself does not contain F_1:

$$E \times E = A_1 + A_2 + E,$$

we obtain a very simple proof of the well-known fact that the twofold degenerate vibration ν_2 does not possess a vibrational angular momentum.

We further see at once that the totally symmetrical vibration q_{1a} cannot enter into any Coriolis terms of the molecule.

Detailed applications to other molecules can be carried out in the same way and will be published elsewhere.

REFERENCES

Childs, W. H. J. 1936 *Proc. Roy. Soc.* A, **153**, 555, and new measurements with increased path length.
Eckart, C. 1935 *Phys. Rev.* **47**, 552.
Jahn, H. A. 1938 *Proc. Roy. Soc.* A, **168**, 469, 495.
Jahn and Childs 1938 *Nature, Lond.*, **141**, 916.
Maue, A.-W. 1937 *Ann. Phys., Lpz.*, **30**, 555.
Nielsen and Nielsen 1935 *Phys. Rev.* **48**, 864.
Tisza, L. 1933 *Z. Phys.* **82**, 48.
Wilson, E. Bright 1935 *J. Chem. Phys.* **3**, 276.

CORRECTION

In Table IV, p. 493 of Part I (*Proc. Roy. Soc.* 1938 A, **168**, 469), in the expression for $(9_{10})_a$ read $\sqrt{(11.17)}$ instead of $\sqrt{77}$.

The emission band spectrum of chlorine. III. Continua

BY THE LATE W. H. B. CAMERON, M.SC. AND A. ELLIOTT, PH.D., D.SC.
Physics Department, University of Sheffield

(*Communicated by S. R. Milner, F.R.S.—Received 5 September* 1938)

[Plate 29]

INTRODUCTION

The work to be described is a continuation of our previous research on the molecular spectra of chlorine (Elliott and Cameron 1937, 1938), and deals with the ultra-violet continua. These continua were described in our earlier paper and the wave-lengths of the maxima of intensity given, for the spectrum excited by the high-frequency electrical discharge in chlorine. From the work of Strutt and Fowler (1912) it was known that the spectrum of chlorine excited by active nitrogen differed from that given by the high frequency discharge. It seemed worth while to re-examine the chlorine afterglow spectrum and to compare it with the high-frequency spectrum; the results, as will be shown, enable a beginning to be made in the interpretation of these continua.

EXPERIMENTAL

Active nitrogen was prepared in the usual way from cylinder nitrogen. The cylinder nitrogen was first stored for some time in large glass vessels in contact with moist phosphorus to remove oxygen, and was then drawn as required through a very extensive drying system, consisting of a tube of soda lime followed by six tubes of phosphorus pentoxide. The total length of the latter tubes was 7 m. The purified nitrogen passed into a discharge tube provided with tungsten electrodes, between which a heavy condensed spark discharge was passed, forming the active modification of nitrogen. This active nitrogen passed into an after-glow tube, and thence over soda lime to a large capacity pump. The chlorine was stored as liquid in a bulb cooled by solid CO_2 and could be allowed to flow into the after-glow tube through a suitably chosen capillary. The spectrum of the after-glow was photographed on Ilford Double X-Press plates with a Hilger medium quartz spectrograph.

It was found necessary to use a considerable quantity of chlorine to produce a spectrum of sufficient intensity; about 1 c.c. of liquid chlorine was employed per hour. No visible glow due to the admission of chlorine was observed, but under the best conditions of working the yellow nitrogen after-glow was not entirely quenched. Some trouble was experienced on account of stray discharge; this caused the nitrogen second positive system to appear on the spectrograms. The stray discharge was reduced sufficiently by introducing an earthed electrode between the discharge and after-glow tubes, and by working with a fairly high gas pressure (16 mm. of mercury). Exposures up to 2 hr. were given.

The spectrum of chlorine excited by active hydrogen may be seen on fig. 2, Plate 29. The chief feature is the continuum whose maximum of intensity is at 2580 A, observed by Strutt and Fowler (1912). The wave-length of this maximum was given as 2564 A in our earlier paper; the difference may be due to a small variation in intensity distribution. The microphotometer record shows that all the maxima of the spectrum excited by a high-frequency discharge are present in the after-glow chlorine spectrum, though with greatly reduced intensity relative to 2580 A. A further and very important difference is that distinct structure is present in the after-glow spectrum on the high-frequency side of the 2580 A maximum; no trace

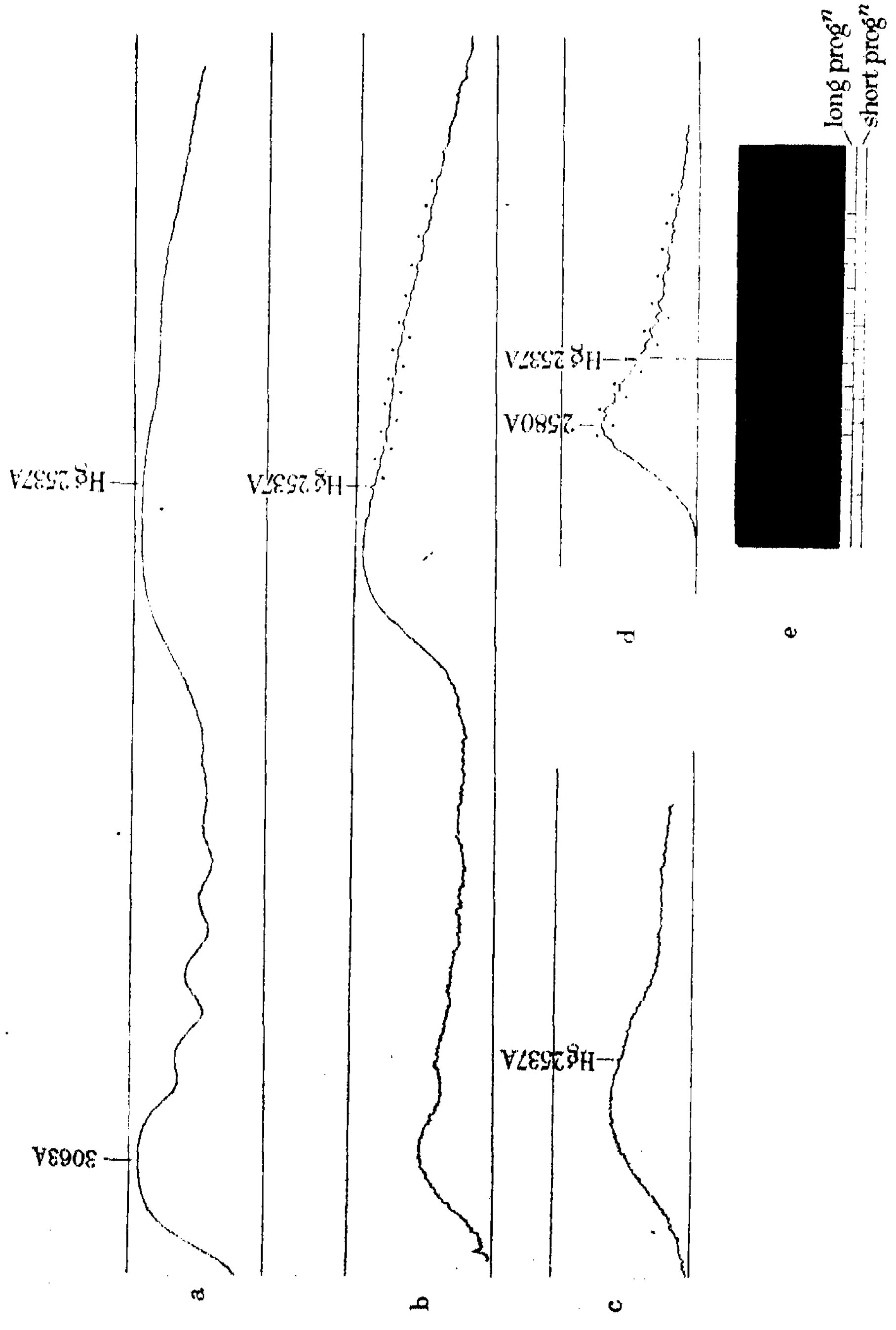


Fig. 2. Emission continua of Cl_2.

of this can be seen on the spectrum excited by a high-frequency discharge. It may, of course, be present but obscured by the other continua which are superposed upon it.* The structure takes the form of a progression of bands, most of them diffuse, and a second shorter progression almost, but not exactly, midway between consecutive members of the first progression. The wave numbers of the maxima of intensity of the bands have been measured and are given in Table I along with the intervals between successive bands. These intervals are about 240 cm.$^{-1}$ and must converge very slowly, for it is impossible to say in which direction they converge; within the rather large experimental error they appear to be constant. Although most of the bands are diffuse, a few at the low-frequency end seem to show some sharpness; this applies as much to one side of a band as to the other. The continuum can be traced at least as far as 2250 A.

Many photographs of the spectrum have been taken under different conditions of pressure, and the effect of cooling the after-glow tube from room temperature down to $-80°$ C has been examined. No change in the intensity distribution within the spectrum was observed on any of the plates. Under some circumstances cooling the tube increased the intensity of the spectrum as a whole, but this appeared to be due primarily to the effect of low temperature on the active nitrogen. The effect was particularly

---

* [*Footnote added* 18 *November* 1938]. It has been suggested that this structure, consisting of diffuse bands, might be attributed to NCl rather than to Cl_2, since it has only been observed in the presence of active nitrogen. Some experiments on the excitation of bromine by active nitrogen, carried out since the above was written, indicate, however, that this is not the case.

The spectrum produced when bromine is admitted into active nitrogen consists of:

(*a*) A band system in the orange region of the spectrum which consists of sharp bands; this system forms the subject of the paper by A. Elliott which follows on p. 469 and is, for reasons stated therein, attributed to NBr.

(*b*) A continuum with maximum of intensity at about 2890 A, which appears at a considerably higher partial pressure of bromine than the bands attributed to NBr. This continuum is doubtless identical in origin with the Br_2 continuum whose maximum of intensity is at 2880 A, obtained in the Tesla discharge in bromine by Ludlam and West (1924). Recent experiments here have shown that this continuum, when excited by active nitrogen, exhibits diffuse bands, similar in character to those obtained with chlorine in active nitrogen, but more closely spaced.

The magnitudes of the nuclear vibration frequencies in NBr ($\omega' = 785$, $\omega'' = 693$ cm.$^{-1}$) are so much greater than the intervals between consecutive bands (ω') in the Br_2 continuum as to make it unlikely that these diffuse bands are due to NBr. The intervals between diffuse bands are approximately constant and equal to 147 cm.$^{-1}$, which is just about what we should expect for the ${}^1\Sigma_u^+$ level in Br_2. For if the vibration frequency in the ${}^1\Sigma_u^+$ state bears the same relation to that in the normal state in Br_2 as in I_2, then ω for the ${}^1\Sigma_u^+$ state in Br_2 would be about 136 cm.$^{-1}$. The interpretation of the diffuse bands attributed to Cl_2 in this paper receives support from these considerations.

TABLE I. WAVE-LENGTHS, WAVE NUMBERS AND INTERVALS OF THE DIFFUSE BANDS IN THE Cl_2 SPECTRUM

Long progression			Short progression		
λ A	Wave number cm.$^{-1}$	Interval cm.$^{-1}$	λ A	Wave number cm.$^{-1}$	Interval cm.$^{-1}$
2581·4	38727	203	2577·3	38788	264
2567·9	38930	214	2559·9	39052	228
2553·9	39144	257	2545·0	39280	246
2537·2	39401	245	2529·2	39526	247
2521·6	39646	239	2513·5	39773	227
2506·4	39885	236	2499·2	40000	222
2491·7	40121	235	2485·4	40222	251
2477·2	40356	245	2470·0	40473	
2462·2	40601	235	2427·9	41176	
2448·1	40836	230			
2434·3	41066	231			
2420·7	41297	255			
2405·9	41552	228			
2392·8	41780	220			
2380·2	42000				

noticeable when the walls of the tube became contaminated with some substance, not yet identified, which quenched the nitrogen after-glow when the tube was at room temperature.

No structure has been observed in any part of the spectrum except that shown in the enlargement on the plate. The continua in the spectrum excited by active nitrogen terminate on the low-frequency side at a wave-length of about 3150 A, but the termination, though steep, is not sharply defined.

INTERPRETATION

Unlike the chlorine emission bands in the visible spectrum (Elliott and Cameron 1938), which are due to the ionized molecule, the continua here considered must be emitted by the neutral chlorine molecule, since active nitrogen, which is used to excite them, is not capable of ionizing chlorine.

The structure in the continuum at 2580 A is certainly vibrational. If the continuum and diffuse bands were due to transitions between a stable and wholly unstable level, only a single progression of diffuse bands would be expected, in addition to the continuum (see for example Kuhn 1930 and Finkelnburg 1938). From the fact that two progressions, if not more, are present, we deduce that both states must be stable over a finite range of internuclear distance. This is also indicated by the fact that some of the bands appear fairly sharp, as if rotation structure might be present. One state must have only a very shallow minimum in its potential energy-internuclear distance curve. Consideration of the Franck-Condon principle together with the fact that the sharpest bands are situated near the position of maximum intensity of the continuum suggests that the stable region of the less stable state (the lower state, in emission spectra of the type considered) lies directly under that part of the upper state which is most populated. Although excitation by active nitrogen is frequently selective, it does not appear to be markedly so here. For if the greatest concentration of molecules were situated in one of the higher vibrational levels of the upper state, we should expect to find two maxima of intensity in the resulting spectrum, in accordance with the Franck-Condon principle. Such is not the case here; none of the other maxima in fig. 2, Plate 29 have the same upper state as 2580 A, as is shown by the fact that the ratio of the intensity of 2580 A to the intensity of the other maxima is dependent on the conditions of excitation. It is not possible to be certain, but it seems likely that the lower vibrational levels of the upper electronic state are the most populated. The main features of the 2580 A continuum and its associated diffuse bands are accounted for qualitatively if we draw the potential energy curves for the levels concerned as in fig. 1. The actual values of the internuclear distances are of course not known.

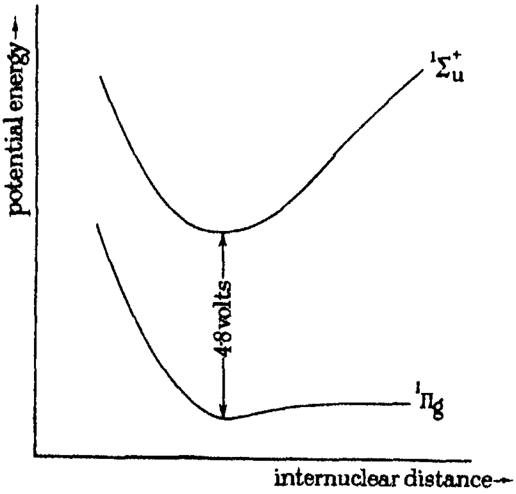


Fig. 1. Qualitative diagram of potential energy curves for Cl_2.

From the fact that the vibrational levels of the upper electronic state converge very slowly, we know that the energy of dissociation of this state must be considerable. At the same time, the vibrational frequency is less than half that for the normal state of chlorine. This unusual combination is also found in the "D" level of iodine, which, according to Mulliken (1934), is a $^1\Sigma_u^+$ state. The correspondence suggests very strongly that the upper electronic level of the 2580 continuum is the corresponding $^1\Sigma_u^+$ state of chlorine. The $^1\Sigma_u^+$ level of iodine is about 6·5 V vertically above the normal state, and it is to be expected that for chlorine the $^1\Sigma_u^+$ level will be somewhat higher. Mulliken (1934) has given reasons for believing that this level (in chlorine) is actually more than 7·9 V (vertically) above the normal state.

With regard to the lower state, we have to find a state of even parity, 39,000 cm.$^{-1}$ or 4·8 V below the $^1\Sigma_u^+$ state and hence more than 3·1 V above the normal state, with a potential energy curve having a shallow minimum. Reference to Table III of Mulliken's paper, together with a consideration of his description of the properties of these states in iodine, makes it probable that the state in question is the $^1\Pi_g$, which in iodine is predicted to be 2·8 V vertically above the ground state.

At this stage little can be said about the other continua. Since they are relatively weak when excited by active nitrogen, they probably come from levels which active nitrogen is just able to excite (that is, with energies relative to the ground state of about 10 V).

SUMMARY

The spectrum of chlorine excited by active nitrogen has been photographed and measured. The continua previously known were recorded, together with a number of diffuse bands, hitherto unknown, associated with the continuum whose maximum of intensity is at a wave-length of 2580 A.

An interpretation of these bands and associated continuum is given, based on the assumption of transitions from a stable upper electronic state, probably analogous to the "D" state of $I_2(^1\Sigma_u^+)$, to a lower electronic state which has only a shallow minimum in its potential energy-internuclear distance curve.

A comparison is made between the chlorine spectrum excited by active nitrogen and by a high-frequency electrical discharge.

REFERENCES

Elliott, A. and Cameron, W. H. B. 1937 *Proc. Roy. Soc.* A, **158**, 681.
— — 1938 *Proc. Roy. Soc.* A, **164**, 531.
Finkelnburg, W. 1938 "Kontinuierliche Spektren", p. 151. Berlin: Julius Springer.
Kuhn, H. 1930 *Z. Phys.* **63**, 458.
Ludlam, E. B. and West, W. 1924 *Proc. Roy. Soc. Edinb.*, **44** (2), 185.
Mulliken, R. S. 1934 *Phys. Rev.* **46**, 549.
Strutt, R. J. and Fowler, A. 1912 *Proc. Roy. Soc.* A, **86**, 105.

DESCRIPTION OF PLATE 29

FIG. 2. (*a*) Microphotometer record of spectrogram showing ultra-violet continua of Cl_2 excited by high-frequency discharge.

(*b*) Record of spectrum of Cl_2 in active nitrogen, on same wave-length scale as (*a*). The diffuse bands are indicated by dots (above the recording line for the long progression, below for the short progression). The bands near the maximum 2580 A cannot be seen, as there the photographic image is too dense. The intensities of the maximum at 2580 are comparable in (*a*) and (*b*).

(*c*) and (*d*) Shorter exposures of the spectra shown in (*a*) and (*b*) respectively, showing more detail near the maximum at 2580 A. Note the absence of structure in the high frequency spectrum (*c*).

(*e*) Enlargement of plate used for making record (*d*), on approximately the same wave-length scale as the record.

A band spectrum attributed to NBr

BY A. ELLIOTT, PH.D., D.SC.
Physics Department, The University of Sheffield

(*Communicated by S. R. Milner, F.R.S.—Received* 26 *September* 1938)

[Plate 30]

INTRODUCTION

In the course of an investigation on the spectra of the halogens and other substances excited by active nitrogen, Strutt and Fowler (1912) observed that, when bromine was admitted into the glowing nitrogen, a feeble orange luminosity was developed. Visual observation with a spectroscope revealed eight narrow bands in the orange region of the spectrum, at intervals of about 28 A. In addition to these bands, a broad symmetrical band with ill-defined edges at 2930 and 2890 A was photographed. This work does not appear to have been followed up. As the apparatus used for a recent similar investigation with chlorine was available (Cameron and Elliott 1938), it seemed worth while attempting to photograph the orange bands with the rapid panchromatic material now available. This attempt has proved successful, and the results will now be described.

EXPERIMENTAL

Liquid bromine was contained in a small bulb immersed in melting ice, connected by means of a long fine capillary tube to the after-glow tube containing streaming active nitrogen. The bromine vapour passed through the capillary tube, and in the after-glow tube the orange glow referred to above was produced. The light from the glowing gas mixture was examined spectroscopically. A single glass prism Littrow instrument, of which the dispersion in the region near 6000 A was 30 A/mm., was used. With Ilford Hypersensitive Panchromatic plates, a good photograph was obtained with an exposure of 6 hr., using a spectroscope slit width of 0·05 mm. An enlargement of this photograph is reproduced in fig. 1, Plate 30. With a slit width of 0·025 mm. and the same exposure time (which was the longest it was found practicable to employ), the stronger bands were recorded, with rather better resolution: this plate was used for the measurement of the vibrational isotope separations (see later). Light from a neon tube gave a comparison spectrum, by reference to which the wave numbers of the band heads were measured. The results are given in Table I, and appear to be correct to one or two wave numbers.

TABLE I. WAVE-LENGTHS AND WAVE NUMBERS OF BAND HEADS ATTRIBUTED TO $^{14}N^{81}Br$

Band	λ A	Wave no. cm.$^{-1}$	Band	λ A	Wave no. cm.$^{-1}$	Band	λ A	Wave no. cm.$^{-1}$
*	6544·4	15276	7–6	6196·0	16135	*	5947·2	16810
6–6	6487·5	15410	8–7	6163·5	16220	8–6	5933·8	16848
*	6468·2	15456	9–8	6127·6	16315	*	5919·7	16888
*	6460·2	15475	2–0	6104·4	16377	9–7	5905·0	16930
7–7	6445·7	15510	3–1	6075·9	16454	10–8	5876·2	17013
*	6418·8	15575	*	6070·0	16470	6–3	5764·4	17343
1–0	6404·8	15610	*	6063·0	16489	7–4	5741·9	17411
2–1	6370·1	15694	4–2	6047·9	16530	8–5	5718·6	17482
9–9	6363·6	15710	*	6037·0	16560	9–6	5695·1	17554
*	6352·7	15737	5–3	6019·5	16608	10–7	5671·5	17627
3–2	6335·4	15780	*	6013·7	16624	7–3	5534·0	18065
4–3	6300·2	15868	*	6007·6	16641	8–4	5515·1	18127
*	6282·4	15913	6–4	5990·7	16688	9–5	5496·6	18188
5–4	6265·9	15955	*	5978·5	16722	10–6	5477·9	18250
6–5	6231·1	16044	7–5	5962·4	16767			

* Indicates that a band has not been classified.

DESCRIPTION OF THE SPECTRUM

Fig. 1, Plate 30 shows that Strutt and Fowler's eight bands form part of a fairly large system of bands, degraded to the violet. The system extends from about 5500 to 6500 A; the latter limit is doubtless due to diminishing sensitivity of the photographic plate. The bands have very well-marked heads, which are quite sharp in the neighbourhood of 6300 A, less so near 5900 A, and from 5700 A towards the violet are clearly double, with about equal intensity in the two heads. The photograph taken when the spectrograph slit was narrow, referred to above, shows doubling of the band heads near 5900 A.

It is worthy of remark that the bands are decidedly stronger in the upper and lower parts of the spectrum than in the middle, showing that the light producing them is stronger near the walls of the after-glow tube than along its axis.

VIBRATIONAL ANALYSIS

The appearance suggests strongly that the spectrum is of the type with well-marked sequences ($v' - v'' =$ constant), rather than that in which progressions are prominent, and the analysis, given in Table II, shows this

TABLE II. VIBRATIONAL SCHEME FOR BAND HEADS ATTRIBUTED TO $^{14}N^{81}Br$

v' \ v''	0		1		2		3		4		5		6		7		8		9
0																			
1	**15610**																		
	767																		
2	16377	683	**15694**																
			760																
3			16454	674	**15780**														
					750														
4					**16530**	662	**15868**												
							740												
5							**16608**	653	15955										
							735		733										
6							17343	655	**16688**	644	16044	634	15410						
							722		723		723								
7							18065	654	17411	644	**16767**	632	16135	625	15510				
									716		715		713						
8									18127	645	17482	634	**16848**	628	16220				
											706		706						
9											18188	634	17554	624	**16930**	615	16315	605	15710
													696		697		698		
10													18250	623	**17627**	614	**17013**		

to be the case. This table, like Table I, gives only the less refrangible of the double heads, when these are resolved.

The absolute values of v'' are decided by the fact that the sequence $v'-v''=1$ terminates abruptly at the 1–0 band; no bands in this sequence can be observed on the low-frequency side of 1–0. No such criterion is available for deciding v', however. For deciding the v' numbering we have two clues, given respectively by the distribution of intensity of the bands in the system and by the vibrational isotope effect. In Table II the most intense bands in a horizontal row (v'' progression) are printed in heavy type. If a smooth curve is drawn through the middle of the squares occupied by these bands, it should, if the numbering is correct, be part of a parabola whose axis is inclined at 45° to the horizontal and passes through the 0–0 band (Franck-Condon parabola). The parabola in the present instance is of the "closed" type, which corresponds to a small difference in the moment of inertia of the molecule in the upper and lower electronic states, and with the numbering adopted satisfies the above conditions. This method of deciding v' is not, however, very accurate. The evidence based on the isotope effect will be considered in the next section. The band head formula to which the heads have been fitted is as follows:

$$\nu_{\text{head}} = 14791 + 785(v'+\tfrac{1}{2}) - 4{\cdot}45(v'+\tfrac{1}{2})^2 - 693(v''+\tfrac{1}{2}) + 5{\cdot}0(v''+\tfrac{1}{2})^2.$$

VIBRATIONAL ISOTOPE EFFECT: EMITTER OF THE SPECTRUM

Since the vibrational frequency of Br_2 in its ground state is 324 cm.$^{-1}$, whereas this constant in the case of the bands under consideration is as high as 785 cm.$^{-1}$, it is evident that these bands cannot be due to Br_2. Under the circumstances it is natural to consider whether the molecule responsible for their production is NBr.

As bromine has isotopes of masses 79 and 81 respectively, present in approximately equal amounts, we should expect all the bands of NBr to consist of two components of approximately equal intensity. It has been shown that the spectrum contains double-headed bands, and these may be ascribed to $^{14}N^{79}Br$ and $^{14}N^{81}Br$. The isotope displacement coefficient $\rho-1$ for these molecules is 0·0019.

Table III gives the measured isotope separations along with those calculated on the assumption that the emitter of the bands is NBr. As the separations were measured on a plate taken when the spectrograph slit was narrow (see above), the results are given to one-tenth of a wave number, though it must be remarked that systematic errors are certain to be present in measurements of band head separations when the heads are only just resolved. The agreement of observed and calculated separations is as good as can be expected under the circumstances; alteration of the v' numbering results in poorer agreement.

TABLE III. SEPARATION OF THE BAND HEADS ATTRIBUTED TO $^{14}N^{79}Br$ AND $^{14}N^{81}Br$, IN cm.$^{-1}$

Band	Obs. separation	Calc. separation
6–4	2·3	3·4
7–5	2·6	3·6
8–6	3·3	3·7
9–7	4·2	3·8
7–4	4·8	4·7
8–5	4·8	4·8
9–6	5·3	4·9

It may be objected that the accuracy of measurement of the isotope effect is insufficient to warrant the assumption that the carrier of the spectrum is NBr and not, for instance, OBr (in which case the isotope displacement coefficient would be 0·0021). Oxygen is known to be a common impurity of active nitrogen, and it is necessary to consider whether OBr can be excluded on chemical grounds.

The apparatus for generating active nitrogen was constructed for the purpose of observing the spectrum of chlorine in the ultra-violet region (Cameron and Elliott 1938). During the earlier stages of this research, the bands of NO were found superposed on the Cl_2 continuum. Special precautions were thereupon taken to eliminate the oxygen responsible for the production of these bands. The quantity of phosphorus in the storage vessels was increased, and its temperature was maintained at well over 20° C, since phosphorus is known to become rather inert at temperatures below about 15° C. A very extensive drying system, consisting in all of about five metres of phosphorus pentoxide tubes, was employed, for Strutt (1917) has shown that water vapour must be removed if the β bands of active nitrogen (NO) are to be suppressed.

These measures were successful; no trace of the NO bands in the Cl_2 spectrum was found even when the exposures were of several hours duration. The same apparatus was used in the experiments described above, and it is certain that the oxygen concentration in the active nitrogen employed was very low. As a further test, a small air leak, in the form of a needle valve connected to the discharge tube, was introduced. The apparatus was run so as to produce the orange glow, and a small quantity of air was allowed to flow into the tube. The orange glow was diminished in intensity, though the yellow after-glow of active nitrogen (due to the first positive bands of N_2) in the space between the discharge tube and the after-glow tube was not at this stage materially affected. Increasing the rate of flow of air into the discharge tube ultimately reduced the intensity of the first positive bands of N_2 as well as that of the bands attributed to NBr. This evidence shows fairly conclusively that oxygen is detrimental to the production of the orange bands, and hence OBr may be excluded as a possible carrier of these bands.

Unclassified bands

Of the bands in Table I which have not been assigned to specific vibrational transitions, some may well be members of fragmentary progressions. Those bands which are situated approximately midway between consecutive members of the sequence in which $v' - v'' = 2$ can hardly be so disposed of, however, for they form a sequence with intervals of about the same magnitude as those in the sequence referred to, and are probably connected with it.

The appearance of these unclassified bands is distinctly different from that of the classified bands which they accompany. In addition to being weaker, their heads are more diffuse and have not that concentration of intensity which characterises the classified bands. A possible explanation is that each strong band is really only a *P* branch, and that the nearest diffuse band on the short wave side comprises the corresponding *R* branch. This would account for the difference in sharpness of the two heads, for the *R* branch, in a band degraded to the violet, does not really form a head in the usual sense of the word, but simply starts from the origin of the band and runs towards the violet with increasing spaces between the lines which comprise it. If this explanation is correct, the rotation constants for the upper and lower levels must differ by a very small amount, for the separation of the band origin and the *P* head (which depends on this difference) is unusually large (of the order of 35 cm.$^{-1}$). An objection which might be raised to the proposed explanation is that the intensity of the *R* branch appears to be much less than that of the *P* branch. This difference is probably only apparent, due to the properties of the photographic plate and to the fact that the concentration of *P* lines near the head is much greater than that of the *R* lines. Because of their apparent low intensity, it is only in the strongest bands that the *R* branches are observable.

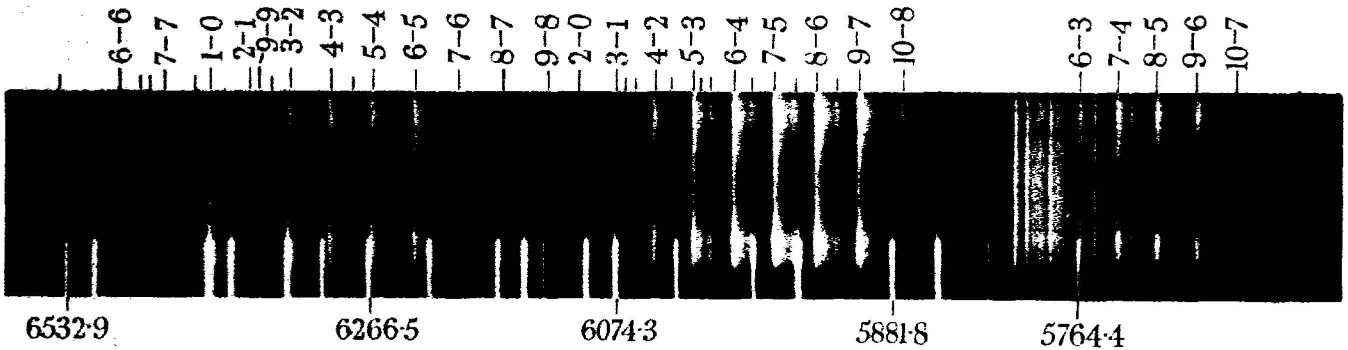


FIG. 1. Band spectrum of NBr.

Upper spectrum: Spectrum of the after-glow of nitrogen containing bromine vapour. The band heads indicated by lines are attributed to NBr; those not so marked are the first positive bands of N_2. Exposure time 6 hr.

Lower spectrum: Comparison spectrum (neon). Exposure time 30 sec.

The spectral region of wave-length below 5656 A has not been included on the plate, as the faintness of the bands there rendered it unsuitable for reproduction.

SUMMARY

1. The spectrum of the after-glow of nitrogen containing bromine vapour has been photographed.

2. The vibrational structure of the band system so observed has been analysed, and the band heads fitted to a formula of the usual type.

3. Considerations based on the isotope effect and on chemical evidence support the view that the bands are emitted by the molecule NBr.

REFERENCES

Cameron, W. H. B. and Elliott, A. 1938 *Proc. Roy. Soc.* A, **169**, 463.
Strutt, R. J. 1917 *Proc. Roy. Soc.* A, **93**, 254.
Strutt, R. J. and Fowler, A. 1912 *Proc. Roy. Soc.* A, **86**, 105.

The general theory of relaxation methods applied to linear systems

By G. Temple

(*Communicated by S. Chapman, F.R.S.—Received* 11 *October* 1938)

1. Introduction and summary

During the last few years Southwell and his fellow-workers have developed a new method for the numerical solution of a very general type of problem in mathematical physics and engineering. The method was originally devised for the determination of stresses in frameworks, but it has proved to be directly applicable to any problem which is reducible to the solution of a system of non-homogeneous, linear, simultaneous algebraic equations in a finite number of unknown variables.* Southwell's "relaxation method" is one of successive approximation and, in order to complete the previous investigations of this method, it is necessary to prove that the successive approximations do actually converge towards the exact solutions. This formal proof is given in § 4.

Southwell's "relaxation" methods are not *directly* applicable to continuous systems, where the number of unknown variables is infinite, but it is shown here that simple extensions and modifications of the relaxation method render it suitable for application to either discrete or continuous systems (§ 5).

The general theory of relaxation methods is then developed in terms of the theory of linear operators (§ 7) and sufficient conditions are prescribed for the convergence of the process of approximation (§ 8).

These general methods are then applied to the solution of non-homogeneous, linear integral equations (§ 10) and to the solution of non-homogeneous, linear differential equations (§§ 11, 12).

Broadly speaking it appears that almost all linear systems of equations, algebraic, integral or differential, can be brought within the scope of relaxation methods, which seem to constitute one of the most powerful methods of computation in mathematical physics and engineering.

---

* In a paper to be published shortly Southwell has extended the method to deal with homogeneous equations such as occur in vibration problems involving characteristic numbers.

I. LINEAR ALGEBRAIC EQUATIONS IN *n* UNKNOWN VARIABLES

2. THE PROBLEM OF STRESS CALCULATION IN FRAMEWORKS

Southwell's method of "Systematic Relaxation of Constraints" was originally devised for the determination of stresses in redundant frameworks, and, although the method appears to be applicable to any minimal problem, it is convenient to retain the notation and terminology of the stress problem. We therefore begin by considering an elastic framework whose configuration is completely specified by n co-ordinates,

$$u_1, u_2, \ldots, u_n,$$

which determine the displacements of the joints of the framework from the positions of equilibrium which they occupy when there are no external forces. In any configuration of the framework which is geometrically possible the generalized forces corresponding to the co-ordinate u_k are

(i) the internal force exerted by the framework, A_k,
(ii) the external force exerted by the loads, B_k, and
(iii) the residual force which would be required to produce equilibrium,

$$C_k = A_k + B_k.$$

The principal properties of an elastic framework are the following:

(i) the internal forces and displacements satisfy a generalized Hooke's Law,
(ii) the framework forms a conservative system, and
(iii) the position of equilibrium under prescribed external forces is stable.

It is upon these properties that relaxation methods depend, and we therefore proceed to express them in analytical form.

(i) In a system which obeys a generalized Hooke's Law the internal force A_k is a homogeneous, linear function of the co-ordinates $\{u_k\}$, i.e.

$$A_k = \Sigma A_{kj} u_j,$$

where the "influence coefficients", A_{kj}, are constants which depend only on the form and the material of the framework, and where the summation here and elsewhere (in I) is with respect to the repeated suffix and is taken over all the joints of the framework. The fundamental problem is to determine the co-ordinates when the framework is in equilibrium under the action of prescribed external forces. In these circumstances there are no residual forces and the co-ordinates satisfy the n simultaneous, linear, algebraic equations,

$$\Sigma A_{kj} u_j + B_k = 0. \tag{2·1}$$

In order that these equations should possess a unique solution for any system of external forces it is necessary and sufficient that the determinant A formed with the influence coefficients should not be zero. We shall denote by A^{kj} the cofactor of A_{kj} in this determinant.

(ii) Since an elastic framework forms a conservative system the influence coefficients satisfy Maxwell's reciprocal relations,

$$A_{kj} = A_{jk},$$

and hence there exists a strain energy function

$$U(u) = \tfrac{1}{2}\Sigma A_{kj} u_k u_j, \tag{2·2}$$

such that the internal forces are given by

$$A_k = \partial U/\partial u_k. \tag{2·3}$$

The potential energy of the external forces is

$$V(u) = \Sigma B_k u_k, \tag{2·4}$$

and the total energy of the framework is

$$W(u) = U(u) + V(u). \tag{2·5}$$

(iii) Finally, since the position of equilibrium is stable, the strain energy $U(u)$ is necessarily positive for all sets of co-ordinates $\{u_k\}$, except, of course, the null-set, $u_k = 0$, $(k = 1, 2, \ldots, n)$; i.e., in the terminology of the theory of operators, the quadratic form (2·2) is "positive definite".

It follows at once that

$$A_{kk} > 0 \quad (k = 1, 2, \ldots, n); \tag{2·6}$$

and a well-known argument shows that the determinant A is essentially positive. The positive character of the strain energy $U(u)$, and, in particular, the relations (2·6) also show that in the position of equilibrium the total energy $W(u)$ is an absolute minimum, $W_{\min}$.

Lastly, we note that, if by an orthogonal transformation of the type

$$u_k = \Sigma l_{kj} x_j,$$

the strain energy is brought into the form

$$U(u) = \tfrac{1}{2}\Sigma L_k x_k^2,$$

while

$$\Sigma u_k^2 = \Sigma x_k^2,$$

then we must have $L_k > 0$ for $k = 1, 2, \ldots, n$. Hence the set of coefficients, $\{L_k\}$, has a positive minimum, say m, and a finite maximum, μ. We can thus infer the important relation, required later (§ 5),

$$0 < m \leqslant 2U(u)/\Sigma u_k^2 \leqslant \mu, \tag{2·7}$$

for all displacements $\{u_k\}$.

3. THE METHOD OF SUCCESSIVE APPROXIMATION

The solution of the system of simultaneous equations (2·1) which determines the position of equilibrium is excessively laborious save for the very simplest frameworks, and, according to Southwell (1935, p. 58), it is effectively impracticable when n, the number of degrees of freedom, exceeds ten or twelve. On the other hand the direct problem of calculating the residual forces $\{C_k\}$ required to give the framework specified co-ordinates $\{u_k\}$ is extremely simple, the requisite residual forces being given explicitly by the equations

$$C_k(u) = \Sigma A_{kj} u_j + B_k. \tag{3·1}$$

Now the residual forces are given by the relation

$$C_k(u) = \partial W/\partial u_k,$$

(by equations (2·2)–(2·5)), and in the position of equilibrium when W is a minimum,

$$C_k(u) = 0.$$

In any other configuration the magnitudes of the residual forces furnish an estimate of the deviation of this configuration from that of equilibrium; and, moreover, we shall now show that a knowledge of the values of the residual force enables us to prescribe displacements which will reduce the value of the total energy W.

To establish this result we observe that if $\{v_k\}$ is a displacement to be considered in this connexion, then it will also be reasonable to consider the whole family of displacements $\{tv_k\}$, where t is an arbitrary, numerical multiplier. We shall describe any displacement of this family as being of the "type" $\{v_k\}$. Since W is a minimum in the equilibrium configuration the best displacement of type $\{v_k\}$ is that which reduces W by the maximum amount. This displacement is easily determined as follows:

In an obvious notation, the total energy in the configuration specified by the co-ordinates $\{u_k+tv_k\}$ is

$$\left.\begin{aligned} W(u+tv) &= W(u)+W_1t+W_2t^2, \\ \text{where}\qquad W_1 &= \Sigma A_{kj}u_kv_j+\Sigma B_kv_k = \Sigma v_kC_k(u), \\ \text{and}\qquad W_2 &= \tfrac{1}{2}\Sigma A_{kj}v_kv_j = U(v). \end{aligned}\right\} \tag{3·2}$$

Hence $$W(u+tv)-W(u) = (t+\tfrac{1}{2}W_1/W_2)^2W_2-\tfrac{1}{4}W_1^2/W_2.$$

Since W_2 is the strain energy for the co-ordinates $\{v_k\}$, it follows that W_2 is positive. Therefore the best displacement of type $\{v_k\}$ is obtained when

$$t = -\tfrac{1}{2}W_1/W_2, \tag{3·3}$$

and then $$W(u+tv)-W(u) = -\tfrac{1}{4}W_1^2/W_2. \tag{3·4}$$

Hence the displacement $\{tv_k\}$ has certainly not increased the total energy.

Moreover, by (3·1)

$$W_1 = \Sigma v_kC_k(u), \tag{3·5}$$

so that, unless the initial configuration $\{u_k\}$ is one of equilibrium (when $C_k(u)=0$), we can always choose the type of the displacement $\{v_k\}$ so that W_1 does not vanish. Then $W(u+tv)$ will be definitely less than $W(u)$, and the configuration $\{u_k+tv_k\}$ will be nearer to the position of equilibrium than the configuration $\{u_k\}$.

It is clear that, starting from any arbitrary configuration, we can thus determine a sequence of configurations in each of which the total energy is less than in the preceding configuration. The configurations so determined will form successive approximations to the equilibrium position. It only remains to show that the sequence of configurations so determined does actually converge to the equilibrium position.

To establish this result, upon which all relaxation methods depend, we write $(u_1^{(p)}, u_2^{(p)}, \ldots, u_n^{(p)})$ for the co-ordinates in the pth configuration of the sequence, and we denote the corresponding value of the total energy by $W(u^{(p)})$. Then

$$W(u^{(p)}) \geqslant W(u^{(p+1)}) \geqslant W_{\min}.$$

Hence the sequence $\{W(u^{(p)})\}$, $(p=1, 2, \ldots)$ converges to a unique limit. Now, by (3·4) and (3·5)

$$W(u^{(p)})-W(u^{(p+1)}) = \tfrac{1}{4}\{\Sigma v_k^{(p)}C_k(u^{(p)})\}^2/U(v^{(p)}),$$

$\{v_k^{(p)}\}$ being the type of displacement employed in forming the $(p+1)$th configuration of the sequence. Therefore,

$$\{\Sigma v_k^{(p)}C_k(u^{(p)})\}^2/U(v^{(p)}) \to 0, \quad \text{as } p\to\infty. \tag{3·6}$$

We shall consider only those methods of approximations in which all the types of displacement $\{v_k^{(p)}\}$ employed are such that the relation (3·6) implies that

$$C_k(u^{(p)}) \to 0, \text{ as } p \to \infty, \quad \text{for } k = 1, 2, \ldots, n. \tag{3·7}$$

When this condition is satisfied the proof of the convergence of the sequence $\{u_k^{(p)}\}$ is easily completed.

In the notation of § 2, the solution of equations (2·1),

$$\Sigma A_{kj} u_k + B_j = 0$$

is

$$A u_k = -\Sigma A^{kj} B_j.$$

Now, in the pth configuration of the sequence, the co-ordinates $\{u_k^{(p)}\}$ satisfy the equation

$$\Sigma A_{kj} u_k^{(p)} + B_j = C_j(u^{(p)}).$$

Hence

$$A u_k^{(p)} = -\Sigma A^{kj} B_j + \Sigma A^{kj} C_j(u^{(p)}).$$

Therefore, by (3·7),

$$A u_k^{(p)} \to -\Sigma A^{kj} B_j, \text{ as } p \to \infty, \quad \text{for } k = 1, 2, \ldots, n,$$

i.e. the sequence of configurations converges to the equilibrium configuration.

We have therefore shown that the co-ordinates in the position of equilibrium, in which W is a minimum, can be determined by a method of successive approximation in which each step in the process depends upon the "direct" problem of the calculation of the residual forces in terms of prescribed co-ordinates. It is perhaps desirable to emphasize that it is the total energy W and not the strain energy U which converges to an extreme value in this process. In Southwell's first paper (1935, § 7) attention was focused on changes in the strain energy, but it is easy to see from simple examples that no *general* statement can be made about the direction of changes in the strain energy.*

The proof of the convergence of the successive approximations given by Black and Southwell (1938, § 3) is not quite complete as it assumes that the convergence of the sequence $\{W(u^{(p)})\}$ always implies the convergence of the sequence $\{u_k^{(p)}\}$. This assumption is easily verified for Southwell's method of successive relaxation, and an elementary formal proof is given in § 4.

---

* If $n = 1$, write $u_1 = x$, $v_1 = y$, $U(x) = \frac{1}{2}Ax^2$, $V(x) = Bx$. Then

$$W(x+ty) - W(x) = (Axy + By)\,t + \tfrac{1}{2}Ay^2t^2,$$

and the right-hand side is a minimum when $t = -(Ax+B)/(Ay)$. The change in the strain energy is

$$U(x+ty) - U(x) = \tfrac{1}{2}A(2x+ty)\,(ty) = -\tfrac{1}{2}A(x^2 - B^2/A^2),$$

and this can be either positive or negative according to the value of x.

4. THE METHOD OF SUCCESSIVE RELAXATION

The preceding section gives the general theory of all relaxation methods for elastic frameworks. Special methods of approximation can now be obtained by particularizing the types of displacement $\{v_k^{(p)}\}$ to be employed at each stage of the process. There appear to be two general methods—Southwell's method of successive relaxation in which the type of displacement is dependent on the co-ordinate system adopted to describe the configuration of the framework, and the method given in the next section in which the type of displacement is determined by the intrinsic properties of the framework as expressed in the total energy function $W(u)$.

In Southwell's original method (1935) the displacement $\{v_k^{(p)}\}$ affects only one co-ordinate, namely that for which the absolute value of the residual force is a maximum; and the magnitude of the change in this co-ordinate is adjusted so as to reduce the corresponding residual force to zero. The method therefore proceeds by a successive relaxation of the constraints represented by the residual forces. In symbols, if

$$|C_k(u^{(p)})| \leqslant |C_m(u^{(p)})| \quad \text{for } k = 1, 2, \ldots, n, \tag{4·1}$$

then

$$v_k^{(p)} = 0 \quad \text{if } k \neq m,$$

$$\text{or} \quad 1 \quad \text{if } k = m.$$

It is easily verified that the magnitude of the displacement prescribed by Southwell's method is the best displacement of this type, i.e. that which reduces the total energy by the greatest amount. In the notation of equation (3·2) we have for Southwell's method,

$$\begin{aligned} W_1 &= \Sigma A_{kj} u_k^{(p)} v_j^{(p)} + \Sigma B_k v_k^{(p)} \\ &= \Sigma A_{km} u_k^{(p)} + B_m \\ &= C_m(u^{(p)}), \end{aligned}$$

and

$$W_2 = \tfrac{1}{2} \Sigma A_{kj} v_k^{(p)} v_j^{(p)} = \tfrac{1}{2} A_{mm}.$$

Hence, by (3·3),

$$t = -\tfrac{1}{2} W_1 / W_2 = -C_m(u^{(p)}) / A_{mm}, \tag{4·2}$$

and the residual force corresponding to the mth co-ordinate is thus reduced to

$$\begin{aligned} C_m(u^{(p)} + t v^{(p)}) &= C_m(u^{(p)}) + t\, \Sigma A_{mk} v_k^{(p)} \\ &= C_m(u^{(p)}) + t A_{mm} = 0, \quad \text{by (4·2)}, \end{aligned}$$

in agreement with the value of the displacement prescribed by Southwell.

It only remains to verify that this type of displacement satisfies the condition that the convergence of the sequence $\{W(u^{(p)})\}$ implies the convergence of the sequence $\{u_k^{(p)}\}$. We have that

$$\begin{aligned} W(u^{(p)}) - W(u^{(p+1)}) &= \tfrac{1}{4}\{\Sigma v_k^{(p)} C_k(u^{(p)})\}^2 / U(v^{(p)}) \\ &= \tfrac{1}{2}\{C_m(u^{(p)})\}^2 / A_{mm}. \end{aligned} \tag{4·3}$$

Now $A_{mm} > 0$ for $m = 1, 2, \ldots, n$, by (2·6), and

$$|C_k(u^{(p)})| \leqslant |C_m(u^{(p)})|, \quad \text{for } k = 1, 2, \ldots, n, \text{ by } (4{\cdot}1).$$

Hence it follows from (4·3) that

$$C_m(u^{(p)}) \to 0, \quad \text{as } p \to \infty,$$

whence $$C_k(u^{(p)}) \to 0, \text{ as } p \to \infty, \quad \text{for } k = 1, 2, \ldots, n.$$

The condition (3·7) is therefore satisfied, and it follows that the sequence of approximation $\{u_k^{(p)}\}$ actually converges to the co-ordinates in the equilibrium configuration.

5. THE METHOD OF STEEPEST DESCENTS

The second general method of successive approximation is most easily visualized by considering the simple example in which there are only two co-ordinates, say x and y, while

$$U = \tfrac{1}{2}(ax^2 + 2hxy + by^2),$$

and $$V = gx + fy, \quad \text{with} \quad |f| < |g|.$$

Then the components of the residual force are

$$C_1(x, y) = ax + hy + g, \quad C_2(x, y) = hx + by + f.$$

If we regard the total energy, $U + V$, as a third co-ordinate z, the expression for the total energy,

$$z = U + V = \tfrac{1}{2}(ax^2 + 2hxy + by^2 + 2gx + 2fy), \tag{5·1}$$

can be regarded as the equation to a surface in ordinary space. The position of equilibrium, given by $C_1 = 0$, $C_2 = 0$, is stable by hypothesis, and therefore corresponds to the bottom of the valley represented by (5·1).

In Southwell's method of approximation the successive displacements which converge to the valley bottom are alternately in the direction of the axes of x and of y, i.e. their direction is determined by the co-ordinate system employed. In the method to be developed in this section each displacement is in the direction of the line of steepest descent at the point reached by the previous displacement. In symbols, Southwell's type of displacement from the position (0, 0) is

$$x = 1,\ y = 0 \quad (\text{since } |f| < |g|),$$

while the type of displacement in the method of steepest descents is

$$x = C_1(0,0) = g, \quad y = C_2(0,0) = f.$$

A few numerical calculations with this example are sufficient to show that Southwell's method is undoubtedly simpler in practical computation, but the method of steepest descents is theoretically better, and, as we shall see later (II), it is easily extended to continuous systems with an infinite number of degrees of freedom to which Southwell's method cannot be applied *directly*.*

For the general framework with n degrees of freedom the direction of "steepest descent" will be given by a displacement of the type

$$v_k^{(p)} = \partial W(u^{(p)})/\partial u_k^{(p)} = C_k(u^{(p)}). \tag{5·2}$$

The direction of this displacement is that in which the total energy W changes most rapidly. It is independent of the co-ordinate system employed, and is organically connected with the intrinsic properties of the framework as expressed in the form of W. In Southwell's method the displacement affects only one co-ordinate, viz. that for which the magnitude of the residual force is a maximum. *In the method of steepest descents the displacement affects all co-ordinates and affects them in the ratio of their residual forces.*

In the notation of § 3, (3·2),

$$W_1 = \Sigma[C_k(u^{(p)})]^2,$$

and

$$W_2 = \tfrac{1}{2}\Sigma A_{kj} C_k(u^{(p)})\, C_j(u^{(p)}).$$

Hence, by (3·3),

$$t = -\tfrac{1}{2}W_1/W_2 = \frac{-\Sigma[C_k(u^{(p)})]^2}{\Sigma A_{kj} C_k(u^{(p)})\, C_j(u^{(p)})},$$

and, by (3·4),

$$W(u^{(p+1)}) - W(u^{(p)}) = \frac{\tfrac{1}{2}\{\Sigma[C_k(u^{(p)})]^2\}^2}{\Sigma A_{kj} C_k(u^{(p)})\, C_j(u^{(p)})}$$

$$= \tfrac{1}{4}\{\Sigma[c_k^{(p)}]^2\}^2/U(c^{(p)}),$$

on writing $c_k^{(p)}$ for $C_k(u^{(p)})$.

---

* It should be added that Southwell has shown that his method can be adapted to continuous systems if these are first replaced (approximately) by appropriate discrete systems. Thus, for example, differential equations must be replaced by difference equations (Bradfield and Southwell 1937; Atkinson, Bradfield and Southwell 1937).

To verify that this type of displacement satisfies the general condition postulated in § 3, we note that by the inequalities (2·7),

$$0 < m \leqslant \frac{2U(c^{(p)})}{\Sigma [c_k^{(p)}]^2} \leqslant \mu.$$

Hence $$W(u^{(p+1)}) - W(u^{(p)}) \geqslant \frac{1}{2\mu} \Sigma [c_k^{(p)}]^2.$$

Now the sequence $\{W(u^{(p)})\}$, $(p = 1, 2, \ldots)$ converges to a limit. Therefore

$$\Sigma [c_k^{(p)}]^2 \to 0 \quad \text{as} \quad p \to \infty$$

and therefore,

$$C_k(u^{(p)}) = c_k^{(p)} \to 0 \text{ as } p \to \infty, \quad \text{for } k = 1, 2, \ldots, n.$$

Thus the successive approximations determined by the method of steepest descents converge to the equilibrium co-ordinates.

In the terminology introduced by Richards (1937), the type of displacement in the method of steepest descents is a "group displacement", and it is clear that, in theory, it is the best possible displacement, although in practice Southwell's methods will probably be quicker for numerical computation.

6. GYROSTATIC SYSTEMS

The relaxation methods described above are essentially methods of successive approximation for the solution of a system of simultaneous, linear, algebraic equations

$$\Sigma A_{kj} u_j + B_k = 0,$$

in the canonical form in which

$$A_{kj} = A_{jk}, \tag{6·1}$$

and $$\Sigma A_{jk} u_j u_k > 0, \tag{6·2}$$

for all $\{u_k\}$ except $u_k = 0$. It has been shown by Black and Southwell (1938, §§ 16–20) that the relaxation method is easily extended to "gyrostatic" systems of equations in which neither the relation (6·1) nor (6·2) is fulfilled.

In fact *any* system of equations

$$\Sigma P_{ik} u_k + Q_i = 0, \tag{6·3}$$

such that $$\det P_{ik} \neq 0,$$

is exactly equivalent to a system in the canonical form. For, let

$$A_{jk} = \Sigma P_{ij} P_{ik},$$

and $$B_j = \Sigma P_{ij} Q_i. \tag{6·4}$$

Then, it follows from (6·3) that

$$\Sigma A_{jk} u_k + B_j = 0, \tag{6·5}$$

and, conversely, since $\det P_{ik} \neq 0$, the original equations (6·3) can be deduced from (6·5).

Now, it is clear from (6·4) that

$$A_{jk} = A_{kj},$$

and that

$$\begin{aligned}\sum_{j,k} A_{jk} u_j u_k &= \sum_{i,j,k} P_{ij} P_{ik} u_j u_k \\ &= \sum_i \{\sum_j P_{ij} u_j\}^2 \geqslant 0.\end{aligned}$$

The sign of equality holds only if

$$\sum_j P_{ij} u_j = 0 \quad \text{for } i = 1, 2, \ldots, n,$$

and these equations are consistent only if $u_j = 0$, $(j = 1, 2, \ldots, n)$. Hence equations (6·5) are in the canonical form. They are, in fact, the familiar conditions that

$$\sum_i \{\sum_j P_{ij} u_j + Q_i\}^2$$

should be a minimum.

II. LINEAR OPERATIONAL EQUATIONS

7. THE PROBLEM OF LINEAR OPERATIONAL EQUATIONS

So far we have applied relaxation methods only to systems with a finite number of degrees of freedom, but the extension to continuous systems with an infinite number of degrees of freedom is immediate, and the necessary generalizations are discovered at once when the problem is stated in terms of the theory of linear operators. In this part of the paper we shall therefore study the general operational problem and in Parts III and IV apply our results to linear integral equations and linear differential equations.

In Part I we dealt with two classes of entities—sets of co-ordinates $\{u_k\}$ and sets of influence coefficients $\{A_{jk}\}$. In this part we replace these by the more general concepts of vectors $\alpha, \ldots$ in a Hilbert space $\mathfrak{h}$, and linear operators $A, \ldots$ acting in this space. These concepts are defined by the following axioms:

The Hilbert space (an abstract, complex, Euclidean space) is linear, metric and complete:

(i) It is *linear* in the sense that if α, β are vectors belonging to $\mathfrak{h}$, so also is $x\alpha + y\beta$, where x, y are any complex numbers.

(ii) It is a *metric* space in the sense that there exists a function (β, α) for any two vectors, which is a complex number such that

$$(\gamma, x\alpha + y\beta) = x(\gamma, \alpha) + y(\gamma, \beta);$$

(β, α) and (α, β) are conjugate complex numbers; and $(\alpha, \alpha) \geqslant 0$.*

(iii) It is a *complete* space in the sense that, if $\{\alpha_p\}$ is an infinite sequence of vectors with the property that

$$|\alpha_m - \alpha_n| \to 0 \quad \text{as} \quad m, n \to \infty,$$

then there exists a vector α such that

$$|\alpha - \alpha_p| \to 0 \quad \text{as} \quad p \to \infty.$$

$\mathfrak{h}$ may have an *infinite* number of dimensions in the sense that for *any* positive integer n there exists n linearly independent vectors, $\alpha_1, \alpha_2, \ldots, \alpha_n$.

For any operator A and any vector α in $\mathfrak{h}$, there exists a "product" $A\alpha$, which is also a vector in $\mathfrak{h}$. The operators are linear in the sense that

$$A(x\alpha + y\beta) = xA\alpha + yA\beta,$$

for all complex numbers x, y and all vectors α, β in $\mathfrak{h}$. The quadratic forms $\Sigma A_{jk} u_j v_k$ are now replaced by expressions of the form $(\beta, A\alpha)$. The most important classes of these forms are defined as follows:

If $(\beta, A\alpha) = (A\beta, \alpha)$ for all α, β in $\mathfrak{h}$, A is said to be "self-adjoint" (symmetric or Hermitian).

If $|(\alpha, A\alpha)| \leqslant k(\alpha, \alpha)$ for all α in $\mathfrak{h}$, A is said to be "bounded".

If $(\alpha, A\alpha) \geqslant 0$ for all α in $\mathfrak{h}$, the sign of equality holding only if $|\alpha| = 0$, A is said to be "positive definite".

The set of linear equations in canonical form,

$$\Sigma A_{kj} u_j + B_k = 0,$$

is now replaced by the operational equation

$$A\alpha + \beta = 0, \tag{7·1}$$

---

* (β, α) is called the "scalar product" of α by β, and $|\alpha|$ the positive value of $(\alpha, \alpha)^{\frac{1}{2}}$, is called the "length" of α. We make considerable use of the inequality $(\alpha, \beta)(\beta, \alpha) \leqslant (\alpha, \alpha)(\beta, \beta)$, which is derived from the principle that $(\alpha + t\beta, \alpha + t\beta)$ is non-negative for all complex numbers t. The "distance function" for the two vectors α and β is $|\alpha - \beta|$. It is assumed that there is a unique null-vector o, such that $|o| = 0$.

where β is a prescribed vector and α is the unknown vector. The generalization of Maxwell's reciprocal relations is provided by the equation

$$(\beta, A\alpha) = (A\beta, \alpha) \tag{7·2}$$

which implies that the "influence" operator A is self-adjoint. The condition of stability, viz.

$$\Sigma A_{jk} u_j u_k \geqslant 0$$

now becomes $(\alpha, A\alpha) > 0$ unless $|\alpha| = 0$, so that A is a positive definite operator.

There is, however, an important difference between the operational problem and the problem of a finite number of linear, algebraic equations. In the latter case we showed (equation (2·7)) that the positive definite character of the strain energy $U(u)$ implied that

$$0 < m \leqslant 2U(u)/\Sigma u_k^2 \leqslant \mu.$$

But it is not true that, because A is positive definite, then

$$0 < m \leqslant \frac{(\alpha, A\alpha)}{(\alpha, \alpha)} \leqslant \mu, \tag{7·3}$$

where m is positive and μ is finite. In fact m may be zero and μ may be infinite. We shall, however, discuss only the case in which the inequalities (7·3) are satisfied. Then A is bounded and may be described as *strictly positive.**

8. THE METHOD OF STEEPEST DESCENTS

Since vectors in a Hilbert space have infinitely many components it is no longer possible to apply the method to successive relaxation, in which only one component of the vector approximating to α is changed at each stage of the process. The method of steepest descents can, however, be extended at once to the operational problem.

The analogue of the total energy $W(u)$ in the framework problem is now

$$W(\alpha) = \tfrac{1}{2}(\alpha, A\alpha) + \tfrac{1}{2}(\alpha, \beta) + \tfrac{1}{2}(\beta, \alpha). \tag{8·1}$$

From the axioms of the last section it follows that $W(\alpha)$ is a real function of α, and we shall now prove that it has a finite lower bound for all vectors

---

* Since we have assumed that $A\alpha$ exists for all α in $\mathfrak{h}$, the inequalities

$$(\alpha, A\alpha) \leqslant \mu(\alpha, \alpha) \quad \text{and} \quad |A\alpha| \leqslant \mu|\alpha|$$

will always hold if $\mathfrak{h}$ is *separable.*

α when β is prescribed. Since (α,β) and (β,α) are conjugate complex numbers, and A is positive definite, it follows that

$$\{(\alpha,A\alpha)+(\alpha,\beta)\}\{(\alpha,A\alpha)+(\beta,\alpha)\}\geqslant 0,$$

i.e.

$$2(\alpha,A\alpha)\,W(\alpha)+(\alpha,\beta)(\beta,\alpha)\geqslant 0.$$

Now

$$0\leqslant(\alpha,\beta)(\beta,\alpha)\leqslant(\alpha,\alpha)(\beta,\beta),$$

and

$$(\alpha,A\alpha)\geqslant m(\alpha,\alpha).$$

Hence

$$W(\alpha)\geqslant -(\beta,\beta)/m. \tag{8·2}$$

The right-hand side of this inequality is independent of α, so that $W(\alpha)$ has a finite lower bound.

Next we consider the problem of improving any approximation α to the solution of the equation $A\alpha+\beta=0$, by replacing α by $\alpha+t\phi$, where ϕ is a prescribed vector and t an arbitrary real number. We have that

$$W(\alpha+t\phi)=W(\alpha)+W_1t+W_2t^2,$$

where

$$\begin{aligned}W_1&=\tfrac{1}{2}(\alpha,A\phi)+\tfrac{1}{2}(\phi,A\alpha)+\tfrac{1}{2}(\beta,\phi)+\tfrac{1}{2}(\phi,\beta)\\&=\tfrac{1}{2}(\gamma,\phi)+\tfrac{1}{2}(\phi,\gamma),\\ \gamma&=A\alpha+\beta,\end{aligned}$$

and

$$W_2=\tfrac{1}{2}(\phi,A\phi).$$

As before the best value of t, giving the greatest decrease in W, is

$$t=-\tfrac{1}{2}W_1/W_2,$$

and then

$$W(\alpha)-W(\alpha+t\phi)=\tfrac{1}{4}W_1^2/W_2.$$

As in the framework problem we shall take the vector ϕ to be determined by γ which now corresponds to the residual forces. In this case, $(\phi=\gamma)$,

$$W_1=(\gamma,\gamma),$$

and

$$W_2=\tfrac{1}{2}(\gamma,A\gamma).$$

Hence the next approximation derived from α is

$$\alpha+t\phi=\alpha-\frac{(\gamma,\gamma)}{(\gamma,A\gamma)}\gamma,$$

and the decrease in the total energy is

$$W(\alpha)-W(\alpha+t\phi)=\frac{(\gamma,\gamma)^2}{2(\gamma,A\gamma)}.$$

Starting with any arbitrary vector α_1, we therefore construct a sequence of vectors, $\alpha_1, \alpha_2, \ldots$, such that

$$\alpha_{p+1} = \alpha_p - \frac{(\gamma_p, \gamma_p)}{(\gamma_p, A\gamma_p)}\gamma_p, \tag{8·3}$$

where $$\gamma_p = A\alpha_p + \beta. \tag{8·4}$$

Then $$W(\alpha_p) - W(\alpha_{p+1}) = \frac{(\gamma_p, \gamma_p)^2}{2(\gamma_p, A\gamma_p)}. \tag{8·5}$$

It follows that the sequence $\{W(\alpha_p)\}$ is monotonic and decreasing. Now by (8·2) it has a finite lower bound. Hence this sequence must converge to a unique limit. Therefore

$$W(\alpha_p) - W(\alpha_{p+1}) \to 0 \quad \text{as} \quad p \to \infty,$$

i.e. by (8·5) $$\tfrac{1}{2}(\gamma_p, \gamma_p)^2/(\gamma_p, A\gamma_p) \to 0 \quad \text{as} \quad p \to \infty.$$

Now $$(\gamma_p, A\gamma_p) \leqslant \mu(\gamma_p, \gamma_p),$$

whence $$(\gamma_p, \gamma_p) \leqslant \mu(\gamma_p, \gamma_p)^2/(\gamma_p, A\gamma_p).$$

Therefore $$(\gamma_p, \gamma_p) \to 0 \quad \text{as} \quad p \to \infty. \tag{8·6}$$

The sequence $\{\gamma_p\}$ thus converges to the null-vector. To establish the convergence of the sequence $\{\alpha_p\}$ we note that

$$|\gamma_p - \gamma_q| \to 0 \quad \text{as} \quad p, q \to \infty.$$

Now, by (8·4), $$\gamma_p - \gamma_q = A(\alpha_p - \alpha_q). \tag{8·7}$$

Hence $$|A(\alpha_p - \alpha_q)| \to 0 \quad \text{as} \quad p, q \to \infty. \tag{8·8}$$

But $$\begin{aligned} m^2(\phi, \phi)^2 &\leqslant (\phi, A\phi)^2 \\ &\leqslant (\phi, \phi)(A\phi, A\phi). \end{aligned}$$

Therefore $$m\,|\phi| \leqslant |A\phi|$$

and $$m\,|\alpha_p - \alpha_q| \leqslant |A(\alpha_p - \alpha_q)|.$$

Hence, by (8·7), $$|\alpha_p - \alpha_q| \to 0 \quad \text{as} \quad p, q \to \infty.$$

Therefore the sequence $\{\alpha_p\}$ converges to a limit vector α.

It is not necessarily true that the sequence $\{A\alpha_p\}$ should converge to $\{A\alpha\}$, but a sufficient condition for the truth of this statement is clearly that

$$|A\alpha| \leqslant \mu\,|\alpha|, \tag{8·9}$$

—a more stringent form of the condition $(\alpha, A\alpha) \leqslant \mu(\alpha, \alpha)$, assumed in (7·3).

We shall assume that the relation (8·9) is verified. Then

$$|A\alpha+\beta| \leqslant |A\alpha_p+\beta| + |A(\alpha-\alpha_p)|$$

$$\leqslant |\gamma_p| + \mu\,|\alpha-\alpha_p|.$$

Hence $$A\alpha+\beta = 0,$$

i.e. the limit vector α does satisfy the original equation (7·1).

The significance of the restrictions (7·3) and (8·9) upon the operator A can be exhibited by introducing the proper vectors ψ_k and the proper values λ_k of this operator. Then

$$A\psi_k = \lambda_k \psi_k,$$

whence $$(\psi_k, A\psi_k)/(\psi_k, \psi_k) = \lambda_k,$$

and $$|A\psi_k|/|\psi_k| = \lambda_k.$$

The conditions therefore imply that all the proper values λ_k lie in a finite range

$$0 < m \leqslant \lambda_k \leqslant \mu < \infty,$$

and, conversely, it can be shown that this restriction on the proper values is sufficient to ensure that the conditions imposed on A are fulfilled.

9. GYROSTATIC SYSTEMS

Similar methods of solution can be applied to any operational equation,

$$P\alpha+\beta = 0,$$

even if P is not self-adjoint, provided that it satisfies the conditions

$$0 < m(\phi,\phi) \leqslant \tfrac{1}{2}\,|(\phi, P\phi)+(P\phi,\phi)|$$

and $$|P\phi| \leqslant \mu\,|\phi|, \tag{9·1}$$

for all vectors ϕ, m and μ being two positive numbers. We write

$$W(\alpha) = \tfrac{1}{2}\,|P\alpha+\beta|^2 = \tfrac{1}{2}(P\alpha, P\alpha)+\tfrac{1}{2}(P\alpha,\beta)+\tfrac{1}{2}(\beta, P\alpha)+\tfrac{1}{2}(\beta,\beta), \tag{9·2}$$

and define a sequence of approximations by the relations

$$\alpha_{p+1} = \alpha_p - \frac{\tfrac{1}{2}(\gamma_p, P\gamma_p)+\tfrac{1}{2}(P\gamma_p,\gamma_p)}{(P\gamma_p, P\gamma_p)}\,\gamma_p, \tag{9·3}$$

where $$\gamma_p = P\alpha_p+\beta.$$

Then $$W(\alpha_p)-W(\alpha_{p+1}) = \frac{[\tfrac{1}{2}(\gamma_p, P\gamma_p)+\tfrac{1}{2}(P\gamma_p,\gamma_p)]^2}{2(P\gamma_p, P\gamma_p)}. \tag{9·4}$$

It follows from (9·2) that $W(\alpha_p) \geqslant 0$, and from (9·4) that

$$W(\alpha_p) \geqslant W(\alpha_{p+1}).$$

Hence the sequence $\{W(\alpha_p)\}$ converges to a unique limit as $p \to \infty$. Hence

$$W(\alpha_p) - W(\alpha_{p+1}) \to 0 \quad \text{as} \quad p \to \infty.$$

Then, by (9·4) $\dfrac{[(\gamma_p, P\gamma_p) + (P\gamma_p, \gamma_p)]^2}{(P\gamma_p, P\gamma_p)} \to 0 \quad \text{as} \quad p \to \infty.$

Now, by (9·1)

$$\frac{m^2}{\mu^2} \mid \gamma_p \mid^2 = \frac{m^2(\gamma_p, \gamma_p)^2}{\mu^2 \mid \gamma_p \mid^2} \leqslant \frac{\mid (\gamma_p, P\gamma_p) + (P\gamma_p, \gamma_p) \mid^2}{(P\gamma_p, P\gamma_p)}.$$

Hence $\mid \gamma_p \mid \to 0 \quad \text{as} \quad p \to \infty.$

To establish the convergence of the sequence $\{\alpha_p\}$ we note that, since

$$\mid \alpha + t\beta \mid^2 \equiv (\alpha, \alpha) + t(\alpha, \beta) + t(\beta, \alpha) + t^2(\beta, \beta)$$

is non-negative for all real t, therefore

$$[(\alpha, \beta) + (\beta, \alpha)]^2 \leqslant 4(\alpha, \alpha)(\beta, \beta).$$

By using (9·1), we therefore find that

$$\begin{aligned} \mid \phi \mid^4 &\leqslant \mid (\phi, P\phi) + (P\phi, \phi) \mid^2/(4m^2) \\ &\leqslant \mid \phi \mid^2 \mid P\phi \mid^2/m^2. \end{aligned}$$

Hence $\mid \phi \mid \leqslant \mid P\phi \mid / m.$

Therefore

$$\begin{aligned} m \mid \alpha_p - \alpha_q \mid &\leqslant \mid P\alpha_p - P\alpha_q \mid \\ &= \mid \gamma_p - \gamma_q \mid, \quad \text{by (9·3)}. \end{aligned}$$

But the sequence $\{\gamma_p\}$ is convergent. Hence the sequence $\{\alpha_p\}$ is also convergent.

If α is the limit of the sequence $\{\alpha_p\}$ then, as in § 8, $P\alpha$ is the limit of the sequence $\{P\alpha_p\}$, and

$$\begin{aligned} \mid P\alpha + \beta \mid &\leqslant \mid P\alpha_p + \beta \mid + \mid P\alpha - P\alpha_p \mid \\ &\leqslant \mid \gamma_p \mid + \mu \mid \alpha - \alpha_p \mid. \end{aligned}$$

Hence $P\alpha + \beta = 0,$

i.e. α satisfies the equation originally proposed.

III. LINEAR INTEGRAL EQUATIONS

10. FREDHOLM'S EQUATION OF THE FIRST KIND

One of the simplest examples of a Hilbert space $\mathfrak{h}$ is furnished by the set of all real, measurable functions $f(x)$ of a variable x, defined in an interval $a \leqslant x \leqslant b$, and such that the integral

$$\int_a^b f^2(x)\,dx$$

exists as a Lebesgue integral. A function of this set can be regarded as a vector with a continuously infinite number of components, and two such functions must be regarded as equivalent if they are equal almost everywhere in (a, b). The "scalar product" of two functions $\alpha(x)$ and $\beta(x)$ is given by

$$(\beta, \alpha) = (\alpha, \beta) = \int_a^b \alpha(x)\,\beta(x)\,dx.$$

The space of these functions is "complete", for, by a well-known theorem due to F. Riesz, the condition

$$\int_a^b [\alpha_p(x) - \alpha_q(x)]^2\,dx \to 0, \quad \text{as} \quad p, q \to \infty$$

is sufficient to ensure the existence of a limit $\alpha(x)$ such that

$$\int_a^b [\alpha(x) - \alpha_p(x)]^2\,dx \to 0 \quad \text{as} \quad p \to \infty,$$

whence $\alpha_p(x) \to \alpha(x)$ almost everywhere in (a, b).

In this realization of the abstract Hilbert space $\mathfrak{h}$ a linear operator A can be defined by the equation

$$A\alpha(x) = \int_a^b A(x, s)\,\alpha(s)\,ds,$$

where the nucleus $A(x, s)$ is a measurable function defined in the region $a \leqslant x$, $s \leqslant b$. The operational equation (7·1) now takes the form

$$\int_a^b A(x, s)\,\alpha(s)\,ds + \beta(x) = 0, \tag{10·1}$$

where $\beta(x)$ and $A(x, s)$ are prescribed functions and $\alpha(x)$ is unknown. This is an integral equation of Fredholm's type and of the first kind.

The conditions to which we have subjected the operator A, equations (7·3) and (8·9), now imply that, for all functions $\alpha(x)$ considered,

$$\left.\begin{aligned} \int_a^b \alpha(x)\,dx \int_a^b A(x,s)\,\alpha(s)\,ds &\geqslant m\int_a^b \alpha^2(x)\,dx, \\ \text{and}\qquad \int_a^b \left\{\int_a^b A(x,s)\,\alpha(s)\,ds\right\}^2 dx &\leqslant \mu^2 \int_a^b \alpha^2(x)\,dx, \end{aligned}\right\} \tag{10·2}$$

where m is a positive number and μ is finite.

We shall first consider the case when A is self-adjoint. It is easily seen that this implies that the nucleus $A(x,s)$ is symmetric in x and s. In these circumstances the analogue of the total energy is given by

$$W(\alpha) = \tfrac{1}{2}\int_a^b\int_a^b A(x,s)\,\alpha(x)\,\alpha(s)\,dx\,ds + \int_a^b \alpha(x)\,\beta(x)\,dx, \tag{10·3}$$

and the sequence of approximations by the methods of steepest descents is given by the formulae—

$$\left.\begin{aligned} \gamma_p(x) &= \int_a^b A(x,s)\,\alpha_p(s)\,ds + \beta(x), \\ \text{and}\qquad \alpha_{p+1}(x) &= \alpha_p(x) - \frac{\gamma_p(x)\int_a^b \gamma_p^2(s)\,ds}{\int_a^b\int_a^b A(x,s)\,\gamma_p(x)\,\gamma_p(s)\,dx\,ds}. \end{aligned}\right\} \tag{10·4}$$

The general argument of § 8 then shows that the sequence of functions $\{\alpha_p(x)\}$ converges almost everywhere to a function $\alpha(x)$ which satisfies the equation (10·1) almost everywhere.

If A is not self-adjoint the formulae are slightly more complicated. Using the method of § 9, and defining $\gamma_p(x)$ as in (10·4), we now define the sequence of approximations by the formula—

$$\frac{\alpha_{p+1}(x) - \alpha_p(x)}{\gamma_p(x)} = \frac{\int_a^b\int_a^b A(x,s)\,\gamma_p(x)\,\gamma_p(s)\,dx\,ds}{\int_a^b \left\{\int_a^b A(x,s)\,\gamma_p(s)\,ds\right\}^2 dx}. \tag{10·5}$$

The sequence $\{\alpha_p(x)\}$ then converges in the same way as before.

IV. LINEAR DIFFERENTIAL EQUATIONS

11. THE REDUCTION OF A LINEAR DIFFERENTIAL SYSTEM TO AN INFINITE SYSTEM OF LINEAR ALGEBRAIC EQUATIONS

Although the application of the general operational theory of Part II to linear integral equations in Part III is so simple and direct, there are unavoidable difficulties in the application to linear differential equations. These difficulties arise from the fact that, if A is a "formal linear differential operator" (Stone 1932, p. 112) the class of functions $\mathfrak{F}$ to which A can be applied is necessarily restricted by conditions of continuity and differentiability, and, moreover, even if the function α belongs to the class $\mathfrak{F}$, this is by no means necessarily true of the function $A\alpha$. To avoid the difficulties of the direct application of the methods of Part II we shall therefore begin by replacing the proposed differential equation by a system of linear algebraic equations in an infinite number of unknown variables.

We consider a formal linear differential operator A of the form

$$A = \sum_{k=0}^{n} p_{n-k}(x)\frac{d^k}{dx^k}, \quad (a \leqslant x \leqslant b),$$

with which we associate a set of linear boundary operators of the form

$$M_k(f) = \sum_{j=1}^{n} \{a_{kj} f^{(j-1)}(a) + b_{kj} f^{(j-1)}(b)\}, \quad k = 1, 2, \ldots, n,$$

where $f^j(x) = df^j(x)/dx$, and a_{kj}, b_{kj} are numerical constants. The fundamental problem is then to solve the differential equation

$$\left.\begin{array}{c} A\alpha(x) + \beta(x) = 0, \\ \text{subject to the homogeneous boundary conditions} \\ M_k(\alpha) = 0, \end{array}\right\} \tag{11·1}$$

$\beta(x)$ being a prescribed function of x.

We shall assume that the differential system (11·1) satisfies conditions analogous to (7·3), which are most conveniently taken in the form

$$0 < m\int_a^b f^2(x)\,dx \leqslant \int_a^b \{Af(x)\}^2\,dx \leqslant \mu\int_a^b f^2(x)\,dx, \tag{11·2}$$

for all functions $f(x)$ satisfying the boundary conditions

$$M_k(f) = 0,$$

m being a positive number and μ being finite.

To define a space $\mathfrak{F}$ for the operator A we introduce an infinite set of functions $\{\phi_n(x)\}$ with the following properties:

(i) Each function satisfies the boundary conditions

$$M_k\{\phi_n\} = 0, \quad k = 1, 2, \ldots, n;$$

(ii) The functions are orthogonal and normal for the range (a, b),

$$\int_a^b \phi_j(x)\,\phi_k(x)\,dx = \begin{matrix} 0 & \text{if } j \neq k, \\ 1 & \text{if } j = k; \end{matrix}$$

(iii) The set of functions is complete, i.e. if

$$\int_a^b f(x)\,\phi_n(x)\,dx = 0 \quad \text{for } n = 1, 2, \ldots,$$

then $f(x)$ is zero almost everywhere in (a, b);

(iv) For each function $A\phi_n$ exists;

(v) The series $\sum_{n=1}^{\infty} \int_a^b (A\phi_n)^2\,dx$ is convergent.

We then define the space $\mathfrak{F}$ to be the set of functions $f(x)$ which can be represented by series of the form

$$\sum_{n=1}^{\infty} c_n \phi_n(x) \tag{11·3}$$

in such a way that

$$\int_a^b \left\{f(x) - \sum_{n=1}^{p} c_n \phi_n(x)\right\}^2 dx \to 0 \quad \text{as } p \to \infty,$$

and we write

$$f(x) \sim \sum_{n=1}^{\infty} c_n \phi_n(x).$$

The series which represents $f(x)$ will converge to $f(x)$ almost everywhere in (a, b).

As in Part III we shall take the scalar product of any two functions $f(x)$, $g(x)$ to be

$$(f, g) = (g, f) = \int_a^b f(x)\,g(x)\,dx,$$

and we shall write $|f|$ for the positive square root of (f, f).

We require three properties of the representation of functions by series of the form (11·3).* Let the sum of the first p terms in the series representing $f(x)$ be denoted by $f_p(x)$, so that

$$f_p(x) = \sum_{n=1}^{p} c_n \phi_n(x),$$

---

* Hobson (1926, 2, 245–49). The series considered in this section have "strong convergence" with exponent 2.

and $$|f-f_p| \to 0, \quad \text{as } p \to \infty.$$

Then (i) $$|f_p| \to |f|, \quad \text{as } p \to \infty; \tag{11·4}$$

(ii) $$(g, f_p) \to (g, f), \quad \text{as } p \to \infty, \tag{11·5}$$

if g belongs to $\mathfrak{F}$; and

(iii) $$Af(x) \sim \Sigma c_n A\phi_n(x). \tag{11·6}$$

The first and second results are well known. To establish the third result we note that, by (11·2),

$$|Af - Af_p| \leqslant \mu^{\frac{1}{2}} |f - f_p|,$$

whence $$|Af - Af_p| \to 0, \quad \text{as } p \to \infty,$$

i.e. $$\int_a^b \left\{Af(x) - \sum_{n=1}^{p} c_n A\phi_n(x)\right\}^2 dx \to 0, \quad \text{as } p \to \infty.$$

We now look for solutions of the differential system (11·1) of the form

$$\alpha(x) \sim \sum_{n=1}^{\infty} a_n \phi_n(x).$$

Since $$A\alpha(x) + \beta(x) \sim \Sigma a_n A\phi_n(x) + \beta(x),$$

it follows from (11·1) that

$$\int_a^b \left\{\sum_{n=1}^{p} a_n A\phi_n(x) + \beta(x)\right\}^2 dx \to 0 \quad \text{as } p \to \infty. \tag{11·7}$$

We shall write $$A_{jk} = \int_a^b A\phi_j(x)\, A\phi_k(x)\, dx \tag{11·8}$$

and $$b_k = \int_a^b \beta(x)\, A\phi_k(x)\, dx.$$

Then (11·7) becomes

$$\sum_{j,k=1}^{p} A_{jk} a_j a_k + 2 \sum_{k=1}^{p} b_k a_k + \int_a^b \beta^2(x)\, dx \to 0, \quad \text{as } p \to \infty.$$

Thus if we write $$W(\alpha) = \tfrac{1}{2} \sum_{j,k=1}^{\infty} A_{jk} a_j a_k + \sum_{k=1}^{\infty} b_k a_k, \tag{11·9}$$

it follows from (11·4), (11·5) and (11·6) that the quadratic form $W(\alpha)$ attains its minimum value, viz.

$$W(\alpha)_{\min} = -\tfrac{1}{2} \int_a^b \beta^2(x)\, dx,$$

when the constants $a_1, a_2, \ldots$ have the values appropriate to the representation of $\alpha(x)$ in the form $\alpha(x) \sim \sum_{n=1}^{\infty} a_n \phi_n(x)$.

The conditions that $W(\alpha)$ should be a minimum are clearly

$$\sum_{j=1}^{\infty} A_{jk}a_j + b_k = 0. \tag{11·10}$$

The original differential system (11·1) has therefore been replaced by an infinite set of linear algebraic equations in the variables $a_1, a_2, \ldots$.

Finally we show that this set of equations is in canonical form. It is obvious from (11·8) that

$$A_{jk} = A_{kj}.$$

Moreover,
$$\sum_{j,k=1}^{p} A_{jk}a_j a_k = \int_a^b \left\{\sum_{n=1}^{p} a_n A\phi_n(x)\right\}^2 dx$$
$$= \int_a^b \{Af_p(x)\}^2 dx, \quad \text{say.}$$

Now
$$\int_a^b \{Af_p(x)\}^2 dx \geqslant m\int_a^b f_p^2(x)\,dx \quad \text{by (11·2)},$$
$$= m\int_a^b \left\{\sum_{n=1}^{p} a_n \phi_n(x)\right\}^2 dx$$
$$= m\sum_{n=1}^{p} a_n^2,$$

since the functions $\{\phi_n(x)\}$ form an orthonormal system. Hence

$$m\sum_{n=1}^{\infty} a_n^2 \leqslant \sum_{n=1}^{\infty} A_{jk}a_j a_k. \tag{11·11}$$

Also,
$$\sum_{n=1}^{p} A_{jk}a_j a_k = \int_a^b \{Af_p(x)\}^2 dx$$
$$\leqslant \mu\int_a^b f_p(x)\,dx \quad \text{by (11·2)},$$
$$= \mu\sum_{n=1}^{p} a_n^2, \quad \text{as before.}$$

Therefore
$$\sum_{n=1}^{\infty} A_{jk}a_j a_k \leqslant \mu\sum_{n=1}^{\infty} a_n^2. \tag{11·12}$$

12. THE SOLUTION OF THE INFINITE SYSTEM OF EQUATIONS BY RELAXATION METHODS

Since the infinite system of equations (11·10) is in canonical form we can apply the method of steepest descents as given in § 8. The vector space $\mathfrak{h}$ now consists of all sets of numbers,

$$\alpha = (a_1, a_2, \ldots, a_n, \ldots)$$

such that the series $\sum_{n=1}^{\infty} a_n^2$ is convergent. The scalar product of the vector α by the vector

$$\beta = (b_1, b_2, \ldots, b_n, \ldots)$$

is now
$$(\beta, \alpha) = (\alpha, \beta) = \sum_{n=1}^{\infty} a_n b_n = \int_a^b \alpha(x)\,\beta(x)\,dx,$$

since the functions $\{\phi_k(x)\}$ form an orthonormal system. The operator A is specified by the infinite set of numbers A_{jk} as defined in (11·7).

Hence to solve the equations (11·10) we start with any arbitrary vector

$$\alpha_1 = (a_1^{(1)}, a_2^{(1)}, \ldots, a_n^{(1)}, \ldots)$$

and form the sequence of vectors defined by the recurrence relation

$$\alpha_{p+1} - \alpha_p = -\frac{(\gamma_p, \gamma_p)}{(\gamma_p, A\gamma_p)}\gamma_p,$$

where
$$\gamma_p = A\alpha_p + \beta.$$

By reasoning based on the quadratic form $W(\alpha)$ of equation (11·9) we can show, as in § 8, that, as $p \to \infty$,

(i) $\{W(\alpha_p)\}$ converges to a unique limit,

(ii) $|\gamma_p| \to 0$,

(iii) α_p converges to a limit vector α.

It is not necessary to employ the condition (8·9) utilized before, for we can now proceed as follows. Let

$$\alpha(x) \sim \sum_{n=1}^{\infty} a_n \phi_n(x),$$

and
$$\alpha_p(x) \sim \sum_{n=1}^{\infty} a_n^{(p)} \phi_n(x).$$

Now
$$|A\alpha - A\alpha_p| \leqslant \mu^{\frac{1}{2}} |\alpha - \alpha_p|, \quad \text{by (11·2).}$$

Therefore,
$$|A\alpha - A\alpha_p| \to 0 \quad \text{as } p \to \infty,$$

and $A\alpha_p$ converges to $A\alpha$ almost everywhere in (a, b). It now follows as in § 8 that the limit function $\alpha(x)$ must satisfy the equation $A\alpha(x) + \beta(x) = 0$ almost everywhere. Finally since $\alpha(x)$ is represented by a series of which each term, $a_k\phi_k(x)$, satisfies the boundary conditions (11·1), so also must $\alpha(x)$ itself. We have thus obtained a complete solution of the problem.

13. CONCLUSION

We have given a general account of the application of relaxation methods to the solution of linear equations. Southwell's original method for linear algebraic equations in a finite number of unknown variables has been the inspiration of the whole investigation. By replacing his method of successive relaxation by our method of steepest descents we have been able to attack linear operational equations, linear integral equations and linear differential equations. In each case our methods give successive approximations which converge to the accurate solution of the problem. Two outstanding problems must be left for future investigation—the numerical solution of particular problems of physics and engineering, and the extension of the methods to non-linear equations.

REFERENCES

Atkinson, R. J., Bradfield, K. N. E. and Southwell, R. V. 1937 *Rep. Memor. Aero. Res. Comm., Lond.*, p. 1822.
Black, A. N. and Southwell, R. V. 1938 *Proc. Roy. Soc.* A, **164**, 447–67.
Bradfield, K. N. E. and Southwell, R. V. 1937 *Proc. Roy. Soc.* A, **161**, 155–81.
Hobson, E. W. 1926 "Theory of Functions of a Real Variable."
Richards, J. C. 1937 *J. Inst. Civ. Eng.*, **5**, 196–216.
Southwell, R. V. 1935 *Proc. Roy. Soc.* A, **151**, 56–95.
Stone, M. H. 1932 "Linear Transformations in Hilbert Space."

The absorption spectra of carbon dioxide and carbon oxysulphide in the vacuum ultra-violet

By W. C. Price* and (Miss) D. M. Simpson

Physical Chemistry Laboratory, Cambridge

(*Communicated by R. G. W. Norrish, F.R.S.—Received* 21 *October* 1938)

[Plate 31]

The present work follows upon an investigation of the absorption spectrum of carbon disulphide (Price and Simpson 1938) which yielded a spectroscopic value of the ionization potential and certain information about the electronic structure of this molecule. Since the three triatomic molecules CO_2, COS and CS_2 are closely related, similarities might be expected to occur in their spectra which would facilitate and help to confirm any analysis suggested for the spectrum of a particular molecule. This expectation has to some extent been realized. It will be convenient first to discuss the spectrum of each molecule separately, and then subsequently to draw comparisons between them. The experimental details by which the spectra were obtained have been described in previous articles. They were photographed with a grazing incidence vacuum spectrograph, the Lyman continuum being used as the continuous background against which the absorption was observed.

THE ABSORPTION SPECTRUM OF CARBON DIOXIDE

The absorption spectrum of carbon dioxide in the vacuum ultra-violet has previously been investigated by Leifson (1926), Henning (1932) and Rathenau (1934) with varying degrees of success. The new material which will be presented here consists mainly in the discovery of a system of bands in the region 1400–1250 A, and in the identification of the members of the electronic series converging to the first photo-ionization limit, which thus yields a spectroscopic value for the minimum ionization potential of the molecule. The dispersion used in the present work is four times as great as that employed by previous investigators, and the amount of detail observable has been correspondingly increased.

At the highest pressures (about 3 mm. Hg in a path length of 80 cm.)

---

* Fellow of Trinity College.

carbon dioxide starts to absorb below 1700 A (see Plate 31). Here some very diffuse bands appear. At 1620 A they merge into continuous absorption, and apart from some narrow regions of transparency between 1250 and 1170 A this extends to all shorter wave-lengths. As the pressure is reduced to a few tenths of a mm. the continuous absorption below 1620 A weakens, and between 1390 and 1240 A a set of distinct bands appears. These are shaded towards the red. On account of the regular intensity distribution of the bands, i.e. the uniform rise to a maximum ~ 1330 A followed by gradual falling away towards short wave-lengths, it is clear that they must correspond either to a single electronic state or to two states rather close together. Although a first glance at the system suggests a single vibrational progression with a single vibration frequency of from 500 to 600 cm.$^{-1}$, the intervals are too irregular for this to be accepted. The most plausible interpretation we have been able to arrive at is that the bands consist of two progressions starting at 1380 A (72,480 cm.$^{-1}$), and 1368 A (73,100 cm.$^{-1}$), each with frequency intervals of about 1225 cm.$^{-1}$ (varying from 1080 to 1350 cm.$^{-1}$) which probably corresponds to the symmetrical valence frequency ν_1. However, on account of their diffuseness, it is doubtful whether the frequencies of the bands can be determined accurately enough to justify completely any one particular analysis. The system is nevertheless important because, as will be seen later, it represents a transition to one of the lower antibonding orbitals of the molecule. One or two of the bands seem to possess double heads, but the majority are too strongly predissociated for this to become apparent.

Below 1290 A each band splits up into two well-separated components. This splitting probably arises from the fact that for vibrational states built up from many quanta, there is a considerable probability that another combination of different vibrations will have the right symmetry and energy to cause a resonance splitting of the levels. This explanation has already been put forward to explain splitting at the high frequency end of certain long progressions in CS_2 and SO_2 (Price and Simpson 1938).

The strongest bands of carbon dioxide lie within the region 1150–690 A. These still give rise to intense absorption even when pressures of only a hundredth of a mm. are present in the spectrograph. Henning (1932) has shown that, in the lower half of this region (830–690 A), the main bands conform approximately to a doublet Rydberg series converging to an ionization potential at 18·00 ± 0·02 V. The bands in the upper region from 1150 to 830 A have not been previously analysed. The new light thrown upon this system in the course of the present work has enabled the following analysis to be given.

At 1130 and 1120 A there occur two extremely strong bands. These, because of their predominance in intensity over all the neighbouring bands, are certainly vibrationless electronic transitions and represent the transition of a non-bonding electron between atomic-like orbitals. The band which comes next in intensity occurs around 995 A. At fairly low pressures this shows up as two bands close together. Since the first ionization potential of CO_2 is known from electron impact experiments to be about 14 V, a short calculation indicates that this band might contain the second Rydberg series members of the bands at 1130 and 1120 A, the separation of the doublet rapidly diminishing in the expected way as the ionization potential is approached. This was confirmed by the finding of higher members of the series corresponding to the Rydberg formula, calculated by taking as first members the mean of the 1130 and 1120 A bands and as second the centre of the 995 A band. The subsequent members satisfied in appearance, intensity and position the usual requirements necessary in order that they might be classed as members of the same Rydberg series. They are best represented by the following formula:

$$\nu_0^n = 111{,}250 - R/(n+0{\cdot}21)^2, \quad n = 2, 3, 4, \text{etc.} \qquad (1)$$

The degree of accuracy with which this formula holds is shown in Table I. The error involved in the extrapolation to the limit is probably not greater than 200 cm.$^{-1}$. The ionization potential is thus 13·73 ± 0·01 V. It has been shown by the work of Mulliken (1935) that the minimum ionization potential of CO_2 corresponds to the removal of a $2p\pi_0$ electron resulting in the formation of the $^2\Pi_g$ state of CO_2^+. The accuracy of our measurements is not sufficient to distinguish the two limits corresponding to the two components of the doublet state of the molecular ion, but it is significant that near the limit many bands were found with a separation of ~ 150 cm.$^{-1}$

TABLE I. TABLE SHOWING OBSERVED AND CALCULATED FREQUENCIES FOR THE CO_2 BANDS OF SERIES (1)

n	I	λ obs.	ν obs.	ν calc.
2	10	1126·1*	88,800*	88,780
3	7	994·0*	100,600*	100,600
4	5	952·1	105·050	105,058
5	3	933·3	107,150	107,207
6		Obscured		108,405
7	1	916·1	109,160	109,140
8	1	912·1	109,640	109,621

* Centre of doublet taken.

which is close to the expected separation. In fact, even the 1130 and the 1120 A bands seemed to be doublets with about this separation.

The value of the ionization potential of CO_2 obtained from series (1) is in excellent agreement with Mulliken's predicted value of $13{\cdot}72 \pm 0{\cdot}03$ V (Mulliken 1935). This prediction is based upon the identification of the 2900 A emission band (Schmid 1933; Smyth 1931) as being the $^2\Sigma_u^+ \rightarrow {}^2\Pi_g$ transition of CO_2^+ and upon the assumption that Henning's Rydberg series corresponds to ionization to the $^2\Sigma$ state. Hence ionization to the $^2\Pi$ state of CO_2^+ must require $18{\cdot}00 - 4{\cdot}28 = 13{\cdot}72$ V ($4{\cdot}28$ V being the ν_e of the 2900 A system). This agrees with the limit we obtain to within the error involved in the extrapolation of the electronic series. The analysis of the CO_2^+ emission bands also involves the assumption that both the states of CO_2^+ are linear. This postulate receives some support from the fact that we find no transverse vibrations among the excited states which lead up to either the ionization potential at 13·73 or that at 18 V. If in the molecular ion the configuration were a bent one, then it might be expected on the Franck-Condon principle that transverse vibrations would accompany the higher states of excitation. The appearance of these vibrations would be allowed by the selection rules appropriate to the altered symmetry, and, moreover, it would be favoured by the change in position of the nuclei. Their non-appearance justifies the assumption that these particular states of the molecular ion are linear.

Two other series similar to (1) were found. They are weaker and less well developed and can roughly be represented by changing the denominator in (1) to (*a*) $(n+0{\cdot}43)^2 \ldots$, $n=3, 4$, etc., and (*b*) $(n+0{\cdot}03)^2 \ldots$, $n=3, 4$, etc. There is a certain amount of confusion towards the limit because of the fact that the first bands of Henning's series seem to fall in the region around 900 A. Henning has only observed the members from $n=3$ upwards. The $n=2$ members can be predicted to be within the range 940–850 A, and we find much strong absorption corresponding to them in this region. The difficulty in assigning particular wave-lengths to these bands is that they split up in a complex way and, moreover, they are superimposed on the convergence limit of series (1). It should further be remarked that Henning has chosen the wrong component of the doublet for the first member ($n=3$) of his published series. From our plates it is clear that it should be about 130,390 instead of the 132,840 cm.$^{-1}$ which he gives. This makes the Rydberg formula fit much better and also leaves 132,840 free for the shorter wave-length series of which it is obviously a member. As Henning does not give the formula for the latter series though he mentions its existence we give it as $\nu = 145{,}800 - R/(n+0{\cdot}95)^2$, which fits the higher members fairly well. It is also interesting to note that nitrogen is the only other molecule for which the limits of two different molecular Rydberg series occurring in absorption in the vacuum ultra-violet have been linked up with two states appearing in the emission spectra of the molecular ion. The Hopfield series (Hopfield 1930; Takamine, Suga and Tanaka 1938) goes to the upper $^2\Sigma_u^+$ state, and the series recently found by Worley and Jenkins (1938) goes to the lower $^2\Sigma_g^+$ state of N_2^+, the transition between these two states giving rise to the well-known negative bands. It may perhaps be of some significance also that nitrogen occupies a position among diatomic molecules somewhat analogous to the position of carbon dioxide among linear triatomic molecules, namely, that both correspond to closed shell configurations of molecular orbitals.

At low pressures most of the bands lying between 1150 and 900 A appear to be slightly shaded towards long wave-lengths, though in many cases it is difficult to decide this with certainty. The 1130 and 1120 A bands are apparent exceptions to this general behaviour, since at medium pressures they appear to spread a little more to short than to long wave-lengths. However, this may be illusory and really due to the presence of some weak bands lying slightly to the short wave-length side of the main bands. The Henning bands between 850 and 690 A are in most cases shaded towards the red, though here again the shading is not very pronounced. Since highly excited molecular states are expected to have nearly the same characteristics as the state of the molecular ion produced by the complete removal of the excited electron, it is to be expected that both the ionic states at 13·7 and 18·0 V should have slightly greater moments of inertia than that of the ground state. This is in agreement with Mulliken's interpretation of the emission data on CO_2^+.

Apart from the bands of a Rydberg type there occurs in the region 1100–900 A a number of weaker bands which form vibrational progressions. Rathenau (1934) reports three such progressions which he calls the "Starke", the "Schwache" and the "Breite" series. Our results support the "Starke" and the "Schwache" series. However, it is difficult to make exact measurements on these bands, because under high dispersion they show multiple heads and consequently some uncertainty arises in trying to set on the middle of the band. It is not possible to accept Rathenau's "Breite" progression, since some of the strong bands it contains are required as Rydberg series members. The reality of the other two series is supported by the intensity distribution. Their frequency differences (~ 1180 cm.$^{-1}$ for the "Starke" and ~ 1230 cm.$^{-1}$ for the "Schwache") most probably correspond to the totally symmetrical valence frequency of

CO_2 which has a value of 1286 cm.$^{-1}$ in the ground state. The broad high-pressure bands falling within a range of 50 A to the long wave-length side of 1130 A are transitions from vibrating ground states. Because of the difficulty of getting very accurate measurements on them little progress has been made in their classification.

THE ABSORPTION SPECTRUM OF CARBON OXYSULPHIDE†

The only previous investigation of the ultra-violet absorption spectrum of COS has been made by Lochte-Holtgreven and Bawn (1932). Using a quartz spectrograph and the H_2 continuum, they found that with pressures of a few tenths of a mm. in a path length of 50 cm., all wave-lengths below 2400 A right down to the limit of transmission of quartz were absorbed. With higher pressures the onset of the absorption moved towards longer wave-lengths eventually reaching a long wave-length limit around 2550 A. In the present work the absorption in the region 2400–1000 A has been investigated with the vacuum spectrograph. It is found that the continuous absorption mentioned above spreads towards shorter wave-lengths as far as 1600 A. At very low pressures, however (~ 0·01 mm.), it breaks up into several wide diffuse bands apparently involving a number of different electronic transitions. The first, for example, extends from about 2380 to 2150 A, the second from 2120 to 2080 A, the third from 2050 to 1860 A, etc. This continuous absorption weakens and fades out towards 1600 A, where there is a definite break in the absorption. At shorter wave-lengths (~ 1550 A) a new electronic transition appears as a group of some seven diffuse bands. The apparent maxima form a progression with a common difference of about 760 cm.$^{-1}$ (Table II). This is no doubt to be

TABLE II. TABLE GIVING THE FREQUENCIES OF BANDS OF THE DIFFUSE (760 cm.$^{-1}$) PROGRESSION OF CARBON OXYSULPHIDE LYING IN THE RANGE 1565–1410 A

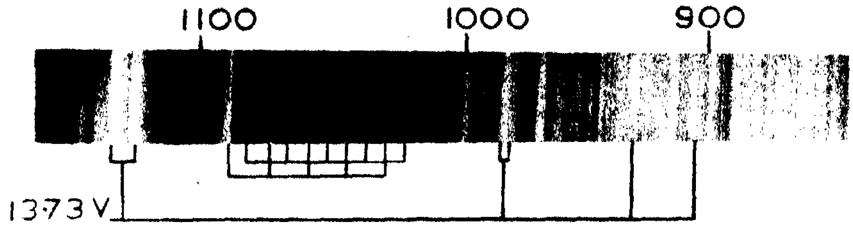
I	ν cm.$^{-1}$	$\Delta\nu$	I	ν cm.$^{-1}$	$\Delta\nu$
2	64,020	—	2	67,070	760
4	64,790	770	3	67,860	790
5	65,580	790	1	68,600	740
5	66,310	730	1	69,370	770

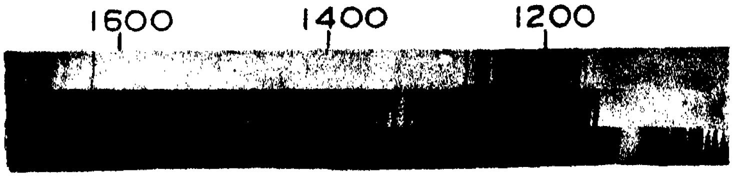
* The frequencies given are estimates of the centres of very wide diffuse bands.

---

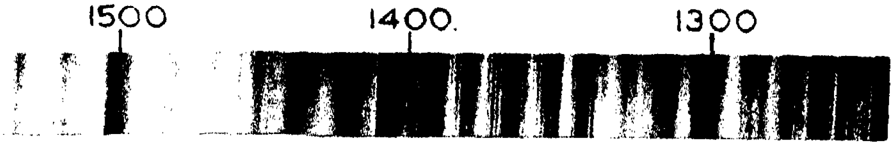
† The name carbon oxysulphide has been used for COS in preference to the alternative carbonyl sulphide because it is clear from thermal data (Pauling and Sherman 1933) that the CO bond is not in the carbonyl state, and that it has a considerable amount of resonance energy with respect to carbonyl CO.

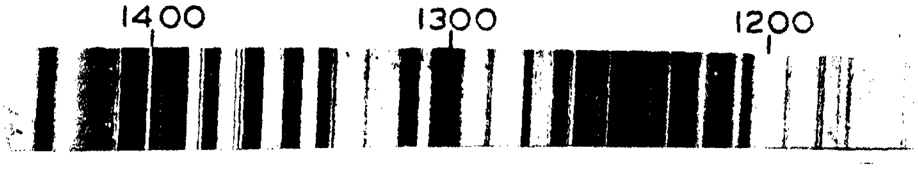
CARBON DIOXIDE





CARBON OXYSULPHIDE





attributed to the ν_1 (symmetrical valence) frequency which has a value of 859 cm.$^{-1}$ in the ground state. Both the selection rules and the fact that as the electron is excited from a closed shell of molecular orbitals, the vibration frequencies are likely to be somewhat diminished by the excitation, support this assignment.

To the short wave-length side of these bands in the range 1410–1215 A there appears a new set of very sharp bands. They are not as strong as the 1550 A bands and require rather higher pressures in order to bring out all their structure.

The individual bands consist of a large number of very sharp components which show the same wide spacing that has proved to be such a puzzling feature of certain similar bands of CS_2. On the low-pressure plates only the strongest components of each band are present. As the pressure increases numerous other weaker components join them. A closer inspection of these bands shows that they fall into two systems: (I) stretching from 1410 to 1310 A, and (II) from 1300 to 1215 A. The first group consists of nine bands, each complex, with from four to nine sharp components. The components themselves show line-like sharpness at low pressures. At higher pressures they are shaded very slightly towards long wave-lengths indicating some slight increase in the OCS distances as a result of the excitation (probably accompanied by some weakening of the OCS bands and diminutions in the associated frequencies). It has been possible to arrange certain of the strong components, which from their appearance seem to be related to one another, into vibrational progressions involving only a single vibrational frequency. Three fairly long progressions have been found throughout the range 1410–1310 A, together with a number of shorter ones among the weaker components (Table III). The

TABLE III. PROGRESSIONS IN THE SHARP 1410–1310 A BAND SYSTEM OF CARBON OXYSULPHIDE

I	ν	$\Delta\nu$	I	ν	$\Delta\nu$	I	ν	$\Delta\nu$
6	70,920	—	8	71,460	—	1	71,552*	—
6	71,640	720	8	72,185	725	6	72,267*	715
6	72,355	715	8	72,900	715	6	72,980	713
6	73,060	705	8	73,610	710	6	73,680	700
5†	73,765	705	8	74,310	700	6†	74,410	730
†	Obscured		7	75,000	690			
5	75,190	712 × 2	7	75,680	680			
2	75,900	710	6	76,350	670			

* Weighted means of a narrow doublet.
† Overlaid with emission lines.

frequency difference is fairly constant the average being ~ 710 cm.$^{-1}$ for the three main series. While the above result shows that a single vibration of around 710 cm.$^{-1}$ is the basis of the progression, it is not put forward as being in any manner a complete analysis. It must be emphasized that the complicated structure of these bands and also similar complexities in many of the bands of CS_2 and CO_2 is something for which there is no analogue in the spectra of diatomic molecules. While it has been suggested in the article on CS_2 that the structure may be rotational in origin due to the molecule becoming *slightly* bent in the excited state there are objections to this explanation. In the case of the bands between 1420 and 1300 A the first band at 1417 A is a single narrow band, the second (~ 1400 A) has two, the third (~ 1385 A) three, and the fourth four predominant components. Weaker additional components come out at higher pressures.

Just below 1300 A another strong vibrational progression is apparent. On the low-pressure plates it appears as three bands, the first being at 1294 A, the second at 1282 A and the third at 1270 A. This main progression has a separation of ~ 715 cm.$^{-1}$, which does not differ greatly from that of the 1420–1300 A bands. However, the intensity distribution in the progression is somewhat different as in this case the first member is the strongest, there being a rapid diminution in intensity in going to the second and third (this is best seen on the low pressure pictures). This intensity distribution is consistent with the correlation of the 715 cm.$^{-1}$ difference with the 859 cm.$^{-1}$ vibration of the ground state indicating as it does a relatively small change in the associated frequency. With increase in pressure many components appear around the main bands. A large number of them can be fitted into short progressions with differences of 710 ± 15 cm.$^{-1}$, and it is fairly sure that a vibration of about this magnitude is the only one involved in the excitation.

TABLE IV. PROGRESSIONS IN THE 1300–1250 A BAND SYSTEM OF CARBON OXYSULPHIDE

I	ν	$\Delta\nu$	I	ν	$\Delta\nu$	I	ν	$\Delta\nu$
10	77,295	—	6	77,420	—	4	77,800	—
7	78,010	715	4	78,150	730	2	78,500	700
4	78,730	720	2	78,865	715	1	79,210	710

Several other weaker progressions of single bands were found in this neighbourhood. These are sufficiently removed from the main progression to be recognizable as different electronic states. They all had the separation of ~ 710 cm.$^{-1}$ and were composed of three or four members slightly shaded towards the red. In all cases the first band was the strongest and the subsequent members rapidly diminished in intensity. The frequencies of the first bands of these progressions together with arbitrary intensities referred to the band at 77,295 cm.$^{-1}$ (1293·7 A) as (100) are 76,830 (5), 77,790 (7), 79,270 (10), 80,830 (5), 81,240 (8). Unlike carbon dioxide and carbon disulphide the spectrum of COS does not contain strong single isolated bands which can be interpreted as transitions between atomic orbitals. The nearest approaches to such transitions are the short vibrational progressions cited above. Possibly the reason for this is that on account of the modified electronic structure resulting from the lack of symmetry (Mulliken 1935), all the excited orbitals, even of relatively non-bonding electrons, have to be a mixture of O, S and C atomic orbitals.

Between about 1200 and 1160 A there appear a set of moderately strong diffuse bands. The difference of ~ 700 cm.$^{-1}$ is again evident, but little can be done with the bands because of this diffuseness. It does appear, however, that the group may form the next Rydberg series member of the set around 1520 A. This would make the ionization potential of COS about 12 V, a value which is intermediate between the values 10·1 V for CS_2 and 13·7 V for CO_2. Extending down from about 1110 A towards shorter wave-lengths, there occur a variety of stronger diffuse bands. These seem to be due to the excitation of a different electron. The wave-lengths at which they appear suggest an ionization potential of ~ 15 V.

Discussion

Mulliken (1935) has already suggested that the first absorption bands of CO_2 (λ max. 1600 A) and CS_2 (λ max. 3200 A) are analogous and correspond to the forbidden transition:

$$ {}^1\Sigma_g^+ \rightarrow (\pi_g)^{-1}\,\sigma,\ {}^1\Pi_g. $$

This is supported by the fact that the absorption maxima of the excited states have similar term values both being ~ 6·2 V (Mulliken's predictions apply to vertical processes and hence the term values to be used are those of the absorption maxima). The bands represent the transition corresponding to the least change of energy, and thus it appears most probable that they involve the excitation of the non-bonding $p\pi$ electrons (which correspond to the minimum ionization potential of the molecule) to some anti-bonding molecular orbital. Particularly good agreement of the term values would not be expected if bonding electrons were involved in the transitions. The diffuse COS absorption below 2500 A (λ max. 2300 A) corresponds roughly to the above-mentioned absorption in CO_2 and CS_2 and occupies an intermediate position. The electronic configuration of this molecule is slightly different on account of the lack of symmetry. As the $g \rightarrow u$ electron rule which makes the transition forbidden in CO_2 and CS_2 does not apply to COS, the absorption is expected to be very much stronger for the unsymmetrical molecule and this is actually found to be so. In the case of CS_2 it will be remembered that the forbidden nature of the 3200 A bands is indicated by the fact that more than a single vibration frequency appears to be present in the excited state, whereas CS_2 possesses only one which is totally symmetrical. The relative weakness of the bands is of course additional evidence. Since for this molecule the longest wave-length bands are discrete, their energy of excitation is probably less than that required for the process $CS_2 \rightarrow CS + S$ (Henri 1931). Only at 2800 A or 4·4 V does predissociation set in. In COS and CO_2 the longest wave-length bands are very diffuse indicating that their energy of excitation already exceeds that necessary for the processes $COS \rightarrow CO + S$ and $CO_2 \rightarrow CO + O$. Thus upper limits of 4·8 V (2550 A) and 7·3 V (1700 A) can be set for the energies necessary to produce such dissociation.

The (λ max. 1335 A) CO_2 bands seem to correspond to the (λ max. 1970 A) CS_2 system, the respective term values of the absorption maxima being 4·5 and 3·8 V. That the CS_2 term value is somewhat less than the CO_2 seems to fit in with what might be expected from a consideration of the atomic term values of O(I) and S(I), e.g. the term values of the ${}^{5,3}S^\circ$ states of O(I) and S(I) which are closest to the term values of the molecular states considered, differ by about 0·6 V, the oxygen term value being the greater (${}^5S^\circ$, $4{\cdot}44 - 3{\cdot}82 = 0{\cdot}62$ V; ${}^3S^\circ$, $4{\cdot}07 - 3{\cdot}48 = 0{\cdot}59$ V). Thus these CO_2 and CS_2 bands are probably to be related to one another and also to the system of COS (λ max. 1510 A) which occupies an intermediate position. The character of the carbon dioxide system together with the moderately high pressures at which it appears suggests that it is to some extent forbidden. The selection rule involved must have little or no effect on the transitions in CS_2 and COS as these are very strong.

The weak CS_2 progression (λ max. 1720 A) does not seem to have any counterpart in the CO_2 spectrum, though it would be difficult to observe since it might be expected to fall in the neighbourhood of the strong 1130–1120 A bands. Carbon oxysulphide, however, appears to have a corresponding system of bands in the set between 1400 and 1300 A. These are obviously "less forbidden" than the CS_2 system, since they come out at moderately low pressures.

Additional analogies on the basis of term values are rather hard to find. The CS_2 band (1595 A) seems possibly related to the strong CO_2 bands at 1130 and 1120 A, the term values being 2·4 and 2·9 V respectively, i.e. the CO_2 term value exceeds that of CS_2 by an amount not very different from the difference of comparable O(I) and S(I) terms; ^5P, $2{\cdot}86 - 2{\cdot}48 = 0{\cdot}38$ and ^3P, $2{\cdot}62 - 2{\cdot}30 = 0{\cdot}32$. However, it is difficult to see why a *single* band in CS_2 should be related to a double band in CO_2.

It should be mentioned also that the series we have found in CO_2 has roughly similar term values to the *AB* series in water (Price 1936), also to the bands below 1350 A in SO_2 (Price and Simpson 1938). These seem possibly related to the $^{5,3}S^\circ$ series of O(I). In CS_2 the bands at 1815 and 1595 A may possibly be related to atomic terms of the type $(^4S)\,4s$, $^3S^\circ$ and $(^4S)\,4p$, ^3P. The molecular term values are about 26,400 and 18,740 as compared with atomic ones of 28,228 and 20,100 cm.$^{-1}$.

In conclusion we wish to thank the Royal Society Grants Committee (W. C. P.) and the Goldsmiths Company (D. M. S.) for financial aid in connexion with this work.

SUMMARY

The absorption spectra of CO_2 and COS have been investigated in the vacuum ultra-violet. In the case of CO_2, the electronic series converging to the lowest $^2\Pi_g$ state of CO_2^+ have been identified and the ionization potential corresponding to this state has been established directly as $13{\cdot}73 \pm 0{\cdot}01$ V. This is in excellent agreement with Mulliken's prediction of $13{\cdot}72 \pm 0{\cdot}03$ V, which is based upon the identification of the 2900 A emission band as being $^2\Sigma_u^+ \rightarrow {}^2\Pi_g$ of CO_2^+, and upon the assumption that the photo-ionization at 685 A is to the $^2\Sigma_u^+$ state. Certain weaker bands not of a Rydberg type have also been found. They are attributed to transitions to anti-bonding molecular orbitals.

The spectrum of COS has been found to be intermediate between that of CO_2 and CS_2. Certain differences are attributed to the lack of symmetry of this molecule. The corresponding electronic states of the three molecules are identified by the criterion of the rough agreement of their molecular term values.

REFERENCES

Henning 1932 *Ann. Phys., Lpz.*, **13**, 599–620.
Henri 1931 *Leipziger Vorträge*, pp. 131–54.
Hopfield 1930 *Phys. Rev.* **36**, 789 (Abstract).
Leifson 1926 *Astrophys. J.* **63**, 73–89.

Lochte-Holtgreven and Bawn 1932 *Trans. Faraday Soc.* **28**, 698–704.
Mulliken 1935 *J. Chem. Phys.* **3**, 720–39.
Pauling and Sherman 1933 *J. Chem. Phys.* **1**, 606–17.
Price 1936 *J. Chem. Phys.* **4**, 147–53.
Price and Simpson 1938 *Proc. Roy. Soc.* A, **165**, 272–91.
Rathenau 1934 *Z. Phys.* **87**, 32–56.
Schmid 1933 *Z. Phys.* **83**, 711–38.
Smyth 1931 *Phys. Rev.* **38**, 2000–15.
Takamine, Saga and Tanaka 1938 *Sci. Pap. Inst. Phys. Chem. Res., Tokyo*, **34**, 854–64.
Worley and Jenkins 1938 *Phys. Rev.* **54**, 305.

Statistical theory of the adsorption of double molecules

By T. S. Chang, *Fitzwilliam House*

(*Communicated by R. H. Fowler, F.R.S.—Received* 24 *October* 1938)

1. INTRODUCTION

The adsorption of gaseous atoms upon a solid surface where the sites of accommodation are regularly arranged has been studied extensively by statistical methods, first without interactions between the adsorbed atoms (Fowler 1935) and later with the interactions properly included. The first approach to include the interactions is made by Fowler, using the Bragg and Williams's approximation (Fowler 1936), and subsequently by Peierls, using Bethe's approximation (Peierls 1936), both methods having been first employed to investigate the order-disorder transitions in alloys. The main results of these calculations are the same. When the forces between the adsorbed atoms are attractive, there are two different types of isotherms separated by a certain temperature called the critical temperature. For isotherms with lower temperature there exists a range of pressure p where several values of the fraction of adsorption θ may correspond to one value of p. Evidently, all these values correspond to stationary values of the free energy, but only one gives the absolute lowest free energy. So when the point for the lowest free energy shifts from one branch $\theta = \theta_1(p)$ to another branch $\theta = \theta_2(p)$, the measured θ will have a discontinuity.

Physically it is as follows. The presence of adsorbed atoms will tend to draw more atoms from the gas to the solid surface, owing to the attractive forces existing between them. After having drawn some atoms to the surface, the attractive forces to draw further atoms are bigger, and more atoms will be drawn. It is possible that when the fraction of adsorption θ is sufficiently large, this process will cumulate so quickly that θ suddenly becomes much larger, behaving like a discontinuity. This process of accumulation will occur only at low temperature, for at high temperatures the rate of evaporation is large, and a statistical equilibrium is reached before any such accumulation takes place.

In some cases an adsorbed particle may need to take up two adjacent sites of accommodation instead of one. An example nearly of this nature is provided by experiments on the competitive adsorption of bivalent copper and univalent sodium on a film of stearate ions on water, where each adsorbed copper ion will displace two adsorbed sodium ions. If there is only one layer of adsorbed copper, say, the mechanism of adsorption will be similar to that of the adsorption of one layer of double molecules B_2 upon a solid surface,

(1) with each molecule B_2 lying completely in the layer,

(2) with each atom B associated with one site of accommodation, and

(3) without dissociation, i.e. the sites of accommodation for two B atoms belonging to the same molecule B_2 are nearest neighbours.

In this paper we shall study the adsorption of gases consisting of B_2 molecules only (i.e. with no single atoms B in the gas) under the above conditions. In actual adsorptions, there is always a partial dissociation on adsorption (if the dissociation is nearly complete as in most actual cases, we can start with the approximation of complete dissociation, see Wang 1937); there are usually several layers of adsorbed molecules; and even when there are only a few adsorbed molecules so that the adsorption approximates to the ideal case of mono-layer adsorption, the adsorbed molecules may not tend to lie completely in the plane of adsorption, but rather to have one end in the plane and the other end higher up above the plane. All these possibilities are contrary to the assumptions (1), (2), and (3) used in our calculations. Nevertheless, it is not without interest to investigate this ideal case owing to its connexion with the stearate films. The interest of competitive adsorption of bivalent and univalent ions on the films was stressed by Dr Langmuir in conversation with Professor Fowler.

The calculations below hold also for adsorption with the adsorbed molecules forming many layers, if we assume that the sites of accommodation for the higher layers, together with those in the bottom layer, form a regular array like a crystal, that each B of the adsorbed molecules is associated with one site, that the sites of accommodation occupied by two

B's of the same molecule are nearest neighbours, and further that the heat of adsorption of a molecule B_2 is the same, whatever its position and orientation may be. These assumptions evidently do not correspond strictly to any actual example.

The main result of the calculations is that if the forces between the adsorbed molecules are attractive, there is also a critical temperature, of the nature described above, and that this critical temperature is higher than that for the adsorption of single atoms, if the interaction potential energy between two B's, situated as nearest neighbours, and belonging to two different molecules, is the same as that between two single atoms, occupying also a pair of nearest neighbours, in the adsorption of single atoms. This means that at some temperature the adsorption of double molecules has a discontinuity, while the adsorption of single atoms has not. As can be seen later, this arises from the various orientations which a double molecule may have, making the probability of adsorption larger than it would otherwise be.

In the course of the calculations, by studying the adsorption of molecules with no forces between them we have obtained the number of ways of arranging $\frac{1}{2}N\theta$ double molecules on N sites. This result may be useful in some other connexions (e.g. regular solutions with double molecules), and will be discussed in detail below.

2. Bethe's approximation

In using statistical methods we have to find the grand partition function of the adsorbed layer (Fowler 1938). Let χ be the energy of desorption of a molecule B_2 when its neighbouring sites are all unoccupied, V the interaction potential energy between two B's belonging to two different molecules and occupying two sites which are nearest neighbours, and let the interaction of two B's belonging to the same molecule be included in χ. Let us write

$$\lambda_g = p_{B_2} h^3 / \{(2\pi m_{B_2})^{\frac{3}{2}} (kT)^{\frac{3}{2}} V_g\}, \tag{1}$$

where p_{B_2} is the pressure in the gaseous phase, m_{B_2} the mass of the molecule B_2, V_g the factor in the partition function of a molecule in the gaseous phase allowing for its rotation and internal oscillations, and h, k, T Planck's constant, Boltzmann's constant, and the absolute temperature respectively. Let V_S be the partition function of an adsorbed molecule occupying two definite sites. V_S therefore contains no part connected with the orientation of the double molecule. We write further

$$\lambda_{B_2} = e^{\chi/kT} \lambda_g V_S, \quad \lambda = e^{\frac{1}{2}V/kT} (\lambda_{B_2})^{\frac{1}{2}}, \quad \eta = e^{-V/kT}, \tag{2}$$

then the grand partition function for the adsorbed layer is

$$\sum_{n,X} f(n,X)\,\lambda^n \eta^X, \tag{3}$$

where n is the number of B's in the layer, X the number of pairs of nearest neighbours BB, including both those belonging to the same molecule and those belonging to different molecules, and $f(n, X)$ the corresponding number of ways of arrangements.

We shall now make use of Bethe's approximation. Let us consider a typical aggregate of sites, with one in the centre, say D_0 for short, and its z nearest neighbours, say $D_1, D_2, \ldots, D_z$. Let us introduce $\theta_0, \theta_1, \theta_2, \ldots, \theta_z$ to denote the state of occupation of $D_0, D_1, \ldots, D_z$, $\theta_0 = 1$ corresponding to D_0 occupied, $\theta_0 = 0$ corresponding to D_0 unoccupied, $\theta_1 = 1$ to D_1 occupied, $\theta_1 = 0$ to D_1 unoccupied, etc. Any configuration of this aggregate can therefore be denoted by $(\theta_0; \theta_1, \theta_2, \ldots, \theta_z)$. The grand partition function for the configuration $(0; \theta_1, \theta_2, \ldots, \theta_z)$ is then

$$\lambda^{(\theta_1+\ldots+\theta_z)}\gamma(\theta_1, \theta_2, \ldots, \theta_z), \tag{4}$$

where the function γ arises from summing over the various configurations of the molecules outside the aggregate. The essence of Bethe's approximation lies in assuming an appropriate form for this function, and determining it afterwards. Here we must be extremely careful, for the result of calculations depends very much upon the approximate function we assume for this γ. In contrast with the adsorption of single atoms, the $\theta_1, \theta_2, \ldots, \theta_z$ are correlated with the molecules outside in two ways:

(1) if any of the neighbours, say D_j, is occupied by a B, the other end of the molecule to which this B belongs must find itself somewhere among the $z-1$ neighbours of D_j outside the aggregate, and

(2) the other $z-2$ nearest neighbours of D_j may be occupied by some B's and interactions exist between the B at D_j and them.

The second effect corresponds to a similar effect in the adsorption of single atoms, i.e. the existence of interactions between adsorbed atoms among the neighbours and the outside, but there is nothing in the adsorption of single atoms parallel to the first. Taking the case in which there is no interaction between the adsorbed particles, and thus leaving (2) out of account, we see that for the adsorption of single atoms, γ will be independent of $\theta_1, \theta_2, \ldots, \theta_z$, whereas for the adsorption of double molecules, due to the correlation (1), the function γ will still depend upon $\theta_1 \ldots \theta_z$. In the following we shall take γ to be

$$c\zeta^{\theta_1+\theta_2+\ldots+\theta_z}, \tag{5}$$

and assume that this form of γ is good enough when both correlations are present. It seems worth while to study first the particular case $V = 0$ to see if the form of γ is good enough when only the first correlation is present.

For this special case, the g.p.f. (short for grand partition function) for $(0; \theta_1, \theta_2, \ldots, \theta_z)$ will be

$$c(\lambda\zeta)^{\theta_1+\theta_2+\ldots+\theta_z}. \tag{6}$$

Next, let us consider the configuration in which the centre is occupied. The other end of the molecule to which the B at D_0 belongs, must find itself at one of the z neighbours, let us say, D_i for definiteness. The g.p.f. for this case will be

$$c\lambda^{2+\theta_1+\ldots+\theta_{i-1}+\theta_{i+1}+\ldots+\theta_z}\gamma_i(\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_z), \tag{7}$$

where γ_i is again a certain function arising from summing all different possible configurations outside the aggregate. By considering the connexion of the neighbours with the outside for this configuration, we see at once that

$$\gamma_i(\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_z) = \gamma(\theta_1, \ldots, \theta_{i-1}, 0, \theta_{i+1}, \ldots, \theta_z), \tag{8}$$

and is, according to our assumption for γ, equal to $c\zeta^{\theta_1+\ldots+\theta_{i-1}+\theta_{i+1}+\ldots+\theta_z}$. In other words, the B and D_i behaves to the outside as an unoccupied site. The complete g.p.f. is therefore given by

$$\sum_{\theta_1\ldots\theta_z} c(\lambda\zeta)^{\theta_1+\ldots+\theta_z} + c\sum_i \sum_{\substack{\theta_1\ldots\theta_{i-1}\\ \theta_{i+1}\ldots\theta_z}} \lambda^2(\lambda\zeta)^{\theta_1+\ldots+\theta_{i-1}+\theta_{i+1}+\ldots+\theta_z}, \tag{9}$$

the first term corresponding to D_0 unoccupied, and the second to D_0 occupied. The summation over "i" in the second sum corresponds to the summation of various orientations, which the molecule whose one end is the B at D_0 may take. The g.p.f. reduces easily to

$$c(1+\epsilon)^z + cz\lambda^2(1+\epsilon)^{z-1}, \tag{10}$$

where $\epsilon = \lambda\zeta$. To determine ζ or ϵ, we employ the condition that the probability for the centre and the neighbours to be occupied is the same, i.e.

The average fraction of adsorption θ

$$= (1/\text{g.p.f.})\{c\sum_i \sum_{\substack{\theta_1\ldots\theta_{i-1}\\ \theta_{i+1}\ldots\theta_z}} \lambda^2(\lambda\zeta)^{\theta_1+\ldots+\theta_{i-1}+\theta_{i+1}+\ldots+\theta_z}\} \tag{11}$$

$$= z^{-1}(1/\text{g.p.f.})\{c\Sigma(\theta_1+\ldots+\theta_z)(\lambda\zeta)^{\theta_1+\ldots+\theta_z}$$
$$+ c\sum_i \sum_{\substack{\theta_1\ldots\theta_{i-1}\\ \theta_{i+1}\ldots\theta_z}} (1+\theta_1+\ldots+\theta_{i-1}+\theta_{i+1}+\ldots+\theta_z)\lambda^2(\lambda\zeta)^{\theta_1+\ldots+\theta_{i-1}+\theta_{i+1}+\ldots+\theta_z}\},$$

which gives

$$z\lambda^2(1+\epsilon)^{z-1} = \epsilon(1+\epsilon)^{z-1} + \lambda^2(1+\epsilon)^{z-1} + (z-1)\lambda^2\epsilon(1+\epsilon)^{z-2}.$$

The final result is

$$\left.\begin{aligned}\lambda^2 &= (z-1)^{-1}\epsilon(1+\epsilon),\\ \theta(1-\theta)^{-1} &= z(z-1)^{-1}\epsilon,\end{aligned}\right\} \tag{12}$$

which, upon eliminating ϵ, gives

$$\lambda^2 = \frac{1}{z^2}\frac{(z-\theta)\,\theta}{(1-\theta)^2}. \tag{13}$$

We have not proved that we can replace $\gamma(\theta_1 \ldots \theta_z)$ in (4) by $c\zeta^{\theta_1+\cdots+\theta_z}$, where γ depends upon $\theta_1, \ldots, \theta_z$ through the fact that corresponding to each adsorbed B among the z neighbours there must be one B belonging to the molecule of this B outside the aggregate, as the centre is unoccupied. The only way of justifying this assumption is to apply the same assumption to a bigger aggregate, where the error due to the inappropriateness of the assumption will be less important, and to compare the results. For this purpose we consider an aggregate consisting of two nearest neighbours D_0, D_0' in the centre, with neighbours $D_2, D_3, \ldots, D_z$, $D_2', D_3', \ldots, D_z'$; $D_0', D_2, D_3, \ldots, D_z$ being the neighbours of D_0, and $D_0, D_2', \ldots, D_z'$ being those of D_0'. Any B among these neighbours may have its other end among the central sites, or outside the aggregate, or among the other neighbours, as the $2z-2$ sites $D_2, \ldots, D_z, D_2', \ldots, D_z'$ may form pairs of nearest neighbours among themselves. We will exclude the third possibility from our calculations, since this involves explicitly the correlations between two neighbours, and will complicate calculations. As before, the grand partition function for any configuration of this aggregate will contain a factor, which arises from summing the various configurations of the molecules outside, and will be assumed to be

$$\bar{c}(\bar{\zeta})^{\Sigma\theta}, \tag{14}$$

where the summation is taken over all B's whose other ends are outside the aggregate, and $\bar{c}$ is a certain constant. c and $\bar{c}$ are different owing to the difference between the sizes of the two different types of aggregates, but their value is of no importance here. The reason of summing over only those B's whose other ends are outside the aggregate is obviously the same as before, i.e. those B's whose other ends are among the central sites behave to the outside as unoccupied sites. The complete grand partition function for our chosen aggregate is found to be

$$\bar{c}(1+\bar{\epsilon})^{2(z-1)} + \bar{c}2(z-1)\,\lambda^2(1+\bar{\epsilon})^{2z-3} + \bar{c}\lambda^2(1+\bar{\epsilon})^{2(z-1)} + \bar{c}(z-1)^2\lambda^4(1+\bar{\epsilon})^{2(z-2)}, \tag{15}$$

where $\bar{\epsilon} = \lambda\bar{\zeta}$. In the above expression, the first term is for the configuration $\theta_0 = \theta_0' = 0$, the second for the configuration $\theta_0 + \theta_0' = 1$, the third for the configuration in which D_0 and D_0' are occupied by a molecule, and the fourth for the configuration in which D_0 and D_0' are occupied by two B's belonging to two different molecules. Using our condition to determine ζ or $\bar{\epsilon}$, we have

$$\begin{aligned}\theta &= (\tfrac{1}{2})\,(\text{g.p.f.})^{-1}\{2\bar{c}(z-1)\,\lambda^2(1+\bar{\epsilon})^{2z-3}+2\bar{c}\lambda^2(1+\bar{\epsilon})^{2(z-1)} \\ &\qquad\qquad +2\bar{c}(z-1)^2\,\lambda^4(1+\bar{\epsilon})^{2(z-2)}\} \\ &= (2z-2)^{-1}\,(\text{g.p.f.})^{-1}\{2\bar{c}(z-1)\,\bar{\epsilon}(1+\bar{\epsilon})^{2z-3}+2\bar{c}(z-1)\,\lambda^2(1+\bar{\epsilon})^{2z-3} \\ &\quad +2\bar{c}(z-1)\,\lambda^2(2z-3)\,\bar{\epsilon}(1+\bar{\epsilon})^{2(z-2)}+2\bar{c}\lambda^2(z-1)\,(1+\bar{\epsilon})^{2z-3} \\ &\quad +2\bar{c}(z-1)^2\,\lambda^4(1+\bar{\epsilon})^{2(z-2)}+\bar{c}(z-1)^2\,(2z-4)\,\lambda^4\bar{\epsilon}(1+\bar{\epsilon})^{2z-5}\},\end{aligned} \tag{16}$$

of which the solution is again (12), with $\bar{\epsilon}$ in place of ϵ. This agreement suffices to show that our assumption for γ will not be seriously wrong.

It is obvious that the above result can be used to find the number of ways of arranging $\tfrac{1}{2}N\theta$ double molecules on N sites. Let the number of configurations of arranging $\tfrac{1}{2}N\theta$ double molecules upon N sites be $g(N,\theta)$. We may safely assume that $g(N,\theta)=e^{Ng^*(\theta)}$, where $g^*(\theta)$ is a function of θ alone, for roughly speaking, if N is doubled and θ is unchanged, $g(N,\theta)$ will be squared. The free energy of the adsorbed phase will be

$$-kT(Ng^*(\theta)+\tfrac{1}{2}N\theta(\chi/kT)+\tfrac{1}{2}N\theta\log V_S), \tag{17}$$

and the free energy of the gaseous phase will be

$$-kTN_g\left\{\log\frac{(2\pi m_{B_2})^{\frac{3}{2}}\,(kT)^{\frac{3}{2}}\,V_g V}{N_g h^3}+1\right\}, \tag{18}$$

where N_g is the number of gaseous molecules and V is the volume of the gas, not to be confused with the interaction energy V. At equilibrium at constant temperature and volume, the free energy is minimum, hence

$$\frac{\partial}{\partial(\frac{1}{2}N\theta)}\{-kT[Ng^*(\theta)+\dots]\}-\frac{\partial}{\partial N_g}\left\{-kTN_g\left[\log\frac{(2\pi m_{B_2}kT)^{\frac{3}{2}}V_g V}{h^3 N_g}+1\right]\right\}=0 \tag{19}$$

or

$$2\frac{dg^*(\theta)}{d\theta}+\log\lambda_{B_2}=0. \tag{20}$$

We can now substitute the value of λ_{B_2} from (13) into the above equation, and obtain

$$2\frac{dg^*(\theta)}{d\theta}+\log\left[\frac{1}{z^2}\frac{(z-\theta)\,\theta}{(1-\theta)^2}\right]=0. \tag{21}$$

Since $g(N,0) = 1$, $g^*(0) = 0$, we have

$$g^*(\theta) = -\frac{1}{2}\int_0^\theta \log\left[\frac{1}{z^2}\frac{(z-\theta)\theta}{(1-\theta)^2}\right]d\theta$$
$$= -\tfrac{1}{2}[z\log z-(z-\theta)\log(z-\theta)+\theta\log\theta+2(1-\theta)\log(1-\theta)-2\theta\log z]. \tag{22}$$

We shall leave discussion of this result to § 5, in connexion with the Bragg and Williams's approximation.

3. BETHE'S APPROXIMATION (CONTINUED)

We consider now the case in which there are interactions between adsorbed molecules. We take as usual an aggregate with one in the centre, and its z nearest neighbours. When the centre is unoccupied, the g.p.f. will be

$$\lambda^{\theta_1+\dots+\theta_z}\gamma(\theta_1,\theta_2,\dots,\theta_z), \tag{23}$$

and we shall assume that

$$\gamma(\theta_1,\theta_2,\dots,\theta_z) = c\zeta^{\theta_1+\theta_2+\dots+\theta_z}. \tag{24}$$

Summing over $\theta_1,\dots,\theta_z$, we get

$$c(1+\lambda\zeta)^z. \tag{25}$$

When the centre and its ith neighbour are occupied by a molecule B_2, the g.p.f. will be

$$\lambda(\eta\lambda)^{1+\theta_1+\theta_2+\dots+\theta_{i-1}+\theta_{i+1}+\dots+\theta_z}\gamma_i(\theta_1,\dots,\theta_{i-1},\theta_{i+1},\dots,\theta_z).$$

It is evident that $\gamma_i(\theta_1,\dots,\theta_{i-1},\theta_{i+1},\dots,\theta_z)$ is no longer $\gamma(\theta_1,\dots,\theta_{i-1},0,\theta_{i+1},\dots,\theta_z)$, because when D_i is occupied it exerts forces on the molecules outside, whereas when D_i is unoccupied no force will be exerted. It is also impossible to take γ_i to be $\gamma(\theta_1,\dots,\theta_{i-1},1,\theta_{i+1},\dots,\theta_z)$, because the B at D_i does not join to another B outside to form a molecule, whereas the other B's among the neighbours do. In other words, the role played by the B at D_i in regard to the relation with the outside is not the same as an unoccupied site, nor the same as the other B's.

To bridge over the trouble we adopt the following procedure. When D_0 and D_1 are occupied by a molecule, we consider the aggregate with D_0 and D_1 in the centre and their $2z-2$ neighbours,* as in the above section, D_1 here taking the place of D_0' there. We exclude as before the possibility for a pair of neighbours to be occupied by a molecule in our calculations, to avoid complications. As D_0 and D_1 are occupied by a molecule, it follows that if any of the neighbours is occupied by a B its other end must be outside the aggregate. Under such conditions the g.p.f. for various states of occupation of the neighbours will be

---

* The assumption of $2z-2$ neighbours for the pair restricts the discussion tacitly to arrays in which the nearest neighbours of a site are not nearest neighbours of each other.

$$\bar{c}\lambda^2\eta(\lambda\eta\bar{\zeta})^{\theta_2+\theta_3+\ldots+\theta_z+\theta_2'+\ldots+\theta_z'}. \tag{26}$$

Summing over $\theta_2 \ldots \theta'_z$, we get

$$\bar{c}\lambda^2\eta(1+\lambda\eta\bar{\zeta})^{2(z-1)}. \tag{27}$$

Similarly, when D_0 and D_1 are both unoccupied, the g.p.f. for various states of occupation of the neighbours will be

$$\bar{c}(\lambda\bar{\zeta})^{\theta_2+\ldots+\theta_z+\theta_2'+\ldots+\theta_z'}. \tag{28}$$

Summing over $\theta_2 \ldots \theta'_z$, we get

$$\bar{c}(1+\lambda\bar{\zeta})^{2(z-1)}. \tag{29}$$

The complete grand partition function is the sum of the expression (25), and z times the expression (27), the first corresponding to D_0 unoccupied, and the second to D_0 occupied. It is therefore

$$c(1+\lambda\zeta)^z+\bar{c}z\lambda^2\eta(1+\lambda\eta\bar{\zeta})^{2(z-1)}. \tag{30}$$

To proceed further, we must get the relation between ζ, $\bar{\zeta}$, c and $\bar{c}$, which can be obtained as follows. Let us take the case in which D_0 and D_1 are both unoccupied. Under such conditions, the g.p.f. for the configuration $(0; 0, \theta_2, \theta_3, \ldots, \theta_z)$ is the expression (23) with θ_1 put to be zero,

i.e. $$c(\lambda\zeta)^{\theta_2+\ldots+\theta_z}.$$

It is also (28) after summing over $\theta'_2, \ldots, \theta'_z$,

i.e. $$\bar{c}(1+\lambda\bar{\zeta})^{z-1}(\lambda\bar{\zeta})^{\theta_2+\ldots+\theta_z}.$$

They should be the same for all $\theta_2, \ldots, \theta_z$.

Hence $$\zeta=\bar{\zeta} \quad \text{and} \quad c=\bar{c}(1+\lambda\zeta)^{z-1}. \tag{31}$$

These are all the relations required.

Making use of the condition that the probability of adsorption is the same at the centre D_0 and its neighbours, we obtain

$$\begin{aligned}\theta &= (\text{g.p.f.})^{-1}\{\bar{c}z\lambda^2\eta(1+\lambda\eta\bar{\zeta})^{2(z-1)}\}\\ &= z^{-1}(\text{g.p.f.})^{-1}\left\{\bar{c}z\lambda^2\eta(1+\lambda\eta\bar{\zeta})^{2(z-1)}+c\zeta\frac{\partial}{\partial\zeta}(1+\lambda\zeta)^z\right.\\ &\qquad\left.+\tfrac{1}{2}\bar{c}\bar{\zeta}\frac{\partial}{\partial\bar{\zeta}}z\lambda^2\eta(1+\lambda\eta\bar{\zeta})^{2(z-1)}\right\},\end{aligned}$$

where care must be taken not to differentiate c and $\bar{c}$.

Using (31) we get
$$\frac{\theta}{1-\theta}=\frac{z}{(z-1)}\frac{\epsilon(1+\eta\epsilon)}{(1+\epsilon)},\tag{32}$$

$$\lambda^2\eta=\frac{\epsilon}{(z-1)}\frac{(1+\epsilon)^{2(z-1)}}{(1+\eta\epsilon)^{2z-3}},\tag{33}$$

giving θ in terms of η and $\lambda^2\eta$ through the parameter ϵ, which is short for $\lambda\zeta$. For $\eta = 1$, (32), (33) reduce to the result above for no interactions.

Before discussing the equations (32), (33), it may be found interesting to deduce a similar formula by kinetic considerations (for kinetic considerations of the adsorption of single atoms and double molecules with complete dissociation on adsorption see J. K. Roberts 1937). The kinetic consideration cannot dispense with the arbitrary quantities like ζ or ϵ, but it makes no use of λ, which is essentially a statistical quantity. First, we consider an aggregate with one central site and its z neighbours. So far as the dependence upon $\theta_1 \dots \theta_z$ is concerned, the probability for the configuration $(0;\theta_1,\theta_2,\dots,\theta_z)$ may be assumed to be proportional to $\phi^{\theta_1+\theta_2+\dots+\theta_z}$, where ϕ is some quantity to be determined later. The probability for one of the neighbours to be occupied when D_0 is unoccupied is therefore $\phi/(1+\phi)$. The probability for $D_2 \dots D_z$ to be occupied when D_0 and D_1 are both unoccupied is also roughly $\phi/(1+\phi)$. Next, we consider the aggregate with D_0, D_1 in the centre, and $2z-2$ neighbours $D_2 \dots D_z$, $D'_2 \dots D'_z$. For this aggregate, the probabilities for various states of occupation of $D_2 \dots D'_z$, with D_0, D_1 unoccupied may be assumed to be proportional to $(\bar{\phi})^{\theta_2+\dots+\theta_z+\theta_2'+\dots+\theta_z'}$, where $\bar{\phi}$ is a quantity similar in nature to ϕ, and then the probability for $D_2 \dots D_z$ to be occupied when D_0 and D_1 are both unoccupied is $\bar{\phi}/(1+\bar{\phi})$. Comparing with the expression $\phi/(1+\phi)$, we get $\phi=\bar{\phi}$. Finally, the probabilities for various states of occupation of $D_2 \dots D'_z$, with D_0, D_1 occupied by a molecule, will be proportional to $(\eta\bar{\phi})^{\theta_2+\dots+\theta_z+\theta_2'+\dots+\theta_z'}$, and the probability for $D_2 \dots D_z$ to be occupied when D_0 and D_1 are occupied by a molecule is $\eta\bar{\phi}/(1+\eta\bar{\phi})$. Our condition gives us at once

$$\theta=(1-\theta)\frac{\phi}{1+\phi}+\theta\left(\frac{1}{z}+\frac{z-1}{z}\frac{\eta\phi}{(1+\eta\phi)}\right)\tag{34}$$

or
$$\frac{\theta}{1-\theta}=\frac{z}{(z-1)}\frac{\phi(1+\eta\phi)}{1+\phi}.\tag{35}$$

This is essentially the same procedure as before, but we have avoided the use of λ. To go further, we shall consider the rate of evaporation and that of condensation. The rate of evaporation of all molecules whose one end is at D_0 is given by

$$A\theta e^{-\chi/kT}\frac{\Sigma\,\eta^{-(\theta_2+\ldots+\theta_z+\theta_2'+\ldots+\theta_z')}\,(\phi\eta)^{\theta_2+\ldots+\theta_z+\theta_2'+\ldots+\theta_z'}}{\Sigma\,(\phi\eta)^{\theta_2+\ldots+\theta_z+\theta_2'+\ldots+\theta_z'}},\tag{36}$$

where A is a certain constant, and the factor $\eta^{-(\theta_2+\ldots+\theta_z')}$ is inserted to take account of the fact that the effective heat of evaporation is decreased or increased by the amount of interactions with the neighbours. The rate of condensation resulting in adsorbed molecules whose one end is at D_0 is

$$\frac{\alpha p_{B_2}}{\mu}(1-\theta)\frac{\Sigma\,(z-\overline{\theta_1+\ldots+\theta_z})\,\phi^{\theta_1+\theta_2+\ldots+\theta_z}}{\Sigma\,\phi^{\theta_1+\theta_2+\ldots+\theta_z}},\tag{37}$$

where $z-\overline{\theta_1+\ldots+\theta_z}$ is the number of unoccupied sites among the z neighbours of D_0, and corresponds to the number of ways of possible adsorptions, α is the condensation coefficient, p_{B_2} the pressure, and μ some function of T, depending upon the rotations and internal oscillations of the gaseous molecules. This reduces immediately to

$$\frac{\alpha p_{B_2}}{\mu}(1-\theta)\,z\,\frac{1}{1+\phi}.\tag{38}$$

Equating (36) and (38), we get

$$\frac{\theta}{1-\theta}=\frac{\alpha p_{B_2}}{\mu}\left(\frac{1}{A}\right)e^{\chi/kT}\frac{z(1+\eta\phi)^{2z-2}}{(1+\phi)^{2z-1}},\tag{39}$$

which, together with (35), gives

$$\frac{\alpha p_{B_2}}{\mu A}e^{\chi/kT}=\frac{1}{(z-1)}\frac{\phi(1+\phi)^{2(z-1)}}{(1+\phi\eta)^{2z-3}}.\tag{40}$$

We see that the formulas (35) and (40) agree with (32), (33) except that $(\alpha p_{B_2}/\mu A)\,e^{\chi/kT}$ has taken the place of $\lambda^2\eta$, which is simply λ_{B_2}, and ϕ the place of ϵ. Of course, we can get the rate of evaporation of molecules at two definite sites, say D_0 and D_1, and also the rate of condensation of molecules upon D_0 and D_1, and equate them, but both are $1/z$ times the corresponding expressions above, so the result is unaffected.

4. ADSORPTION ISOTHERMS AND THE SECOND APPROXIMATION

The equations (32), (33) are plotted below, with θ as abscissa and $\log_{10}\lambda_{B_2}$ as ordinate, for various temperatures, for molecules with attractive forces in fig. 1, and for molecules with repulsive forces in fig. 2. We see that for the case of attractive forces there are two kinds of isotherms, separated by the critical temperature T_c. For those isotherms whose temperature is lower than T_c, the actual θ will trace that branch of the function $\theta = \theta(p)$ from where it comes, until the state of the absolute minimum of the free energy shifts to another branch, whereupon a discontinuity results. The critical temperature can be determined by

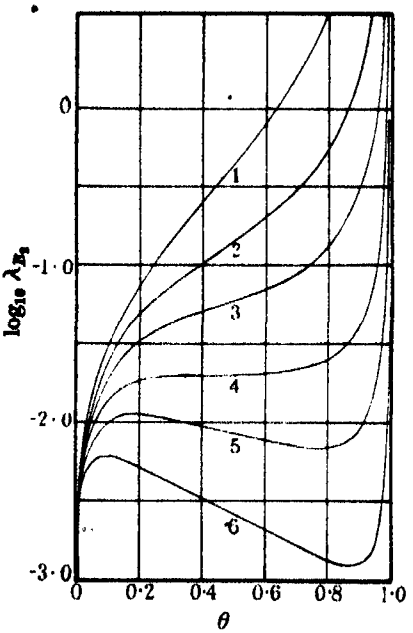


FIG. 1

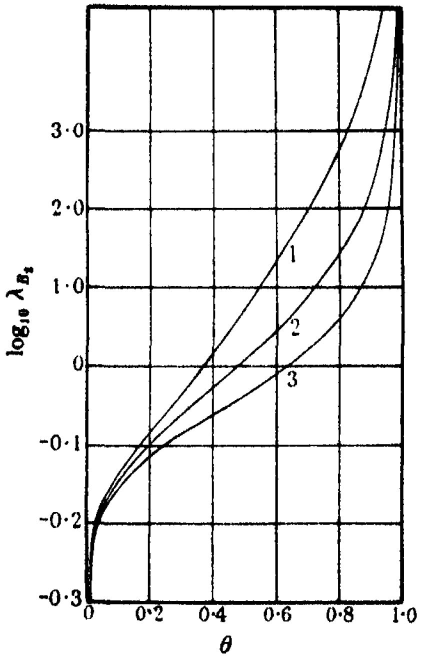


FIG. 2

FIG. 1. Curves showing $\log_{10}\lambda_{B_2}$ against θ for attractive interactions between adsorbed molecules. Curves: (1) $\eta = 1 (T = \infty)$, (2) $\eta = 1{\cdot}50$, (3) $\eta = 2{\cdot}04$, (4) $\eta = 2{\cdot}97 (\eta_c)$, (5) $\eta = 4{\cdot}00$, (6) $\eta = 6{\cdot}00$.

FIG. 2. Curves showing $\log_{10}\lambda_{B_2}$ against θ for repulsive interactions between adsorbed molecules. Curves: (1) $\eta = 0{\cdot}4$, (2) $\eta = 0{\cdot}7$, (3) $\eta = 1{\cdot}0 (T = \infty)$.

$$\frac{d\lambda_{B_2}}{d\theta} = \frac{d^2\lambda_{B_2}}{d\theta^2} = 0. \tag{41}$$

From (32) $$d\theta\left(\frac{1}{\theta} + \frac{1}{1-\theta}\right) = \left(\frac{1}{\epsilon} + \frac{\eta}{1+\eta\epsilon} - \frac{1}{1+\epsilon}\right)d\epsilon,$$

so that $d\theta/d\epsilon \geqslant 0$. Hence (41) is equivalent to

$$d\lambda_{B_2}/d\epsilon = d^2\lambda_{B_2}/d\epsilon^2 = 0.$$

At $d\lambda_{B_2}/d\epsilon = 0$,

$$1/\epsilon + (2z-1)/(1+\epsilon) - (2z-3)\,\eta/(1+\eta\epsilon) = 0.$$

This is a quadratic equation showing that there are either two points, or none, on the isotherm at which $d\lambda_{B_2}/d\theta = 0$. From this equation and $d^2\lambda_{B_2}/d\epsilon^2 = 0$ we find that the critical temperature is given by

$$\eta_c = e^{-V/kT_c} = \tfrac{1}{2}\{(2z^2-5z+4) + 2\,(2z^2-5z+3)^{\frac{1}{2}}\}/(z-2)^2. \tag{42}$$

We give a table below showing the values of η_c and T_c for various values of z, including also the values for the adsorption of single atoms for comparison.

$z=3$		$z=4$		$z=6$		$z=8$		$z\to\infty$	
η_c	T_c	η_c	T_c	η_c	T_c	η_c	T_c	η_c	T_c
Adsorption of single atoms									
9·00	$-\frac{V}{k}0{\cdot}455$	4·00	$-\frac{V}{k}0{\cdot}721$	2·25	$-\frac{V}{k}1{\cdot}23$	1·78	$-\frac{V}{k}1{\cdot}74$	$1+\frac{4}{z}$	$-\frac{V}{k}\frac{z}{4}$
Adsorption of double molecules									
5·95	$-\frac{V}{k}0{\cdot}561$	2·97	$-\frac{V}{k}0{\cdot}919$	1·85	$-\frac{V}{k}1{\cdot}63$	1·54	$-\frac{V}{k}2{\cdot}32$	$1+\frac{3+2\sqrt{2}}{2z}$	$-\frac{V}{k}\frac{2z}{3+2\sqrt{2}}$

It can be seen that the critical temperature here obtained is higher than that for the adsorption of single atoms, if V is the same. We have not plotted the isotherms for the adsorption of single atoms, but from the equations*

$$\xi = \frac{\epsilon(1+\epsilon)^{z-1}}{(1+\eta\epsilon)^{z-1}}, \quad \frac{\theta}{1-\theta} = \frac{\epsilon(1+\eta\epsilon)}{1+\epsilon}, \tag{43}$$

where $$\xi = e^{\chi/kT}\{ph^3V_S/(2\pi m)^{\frac{3}{2}}\,(kT)^{\frac{5}{2}}\,V_g\}$$

(see Peierls 1936), we know that at the points θ and $1-\theta$ the value of $d\log\xi/d\theta$ is the same, and in particular, if the isotherms have stationary values, the maximum is equally far from $\theta = 0$ as the minimum from $\theta = 1$. Here, on the other hand, if θ is fairly small, the value of $(d\log\lambda_{B_2}/d\theta)$ at the point $(1-\theta)$ is larger than that at θ, and the maximum is nearer to $\theta = 0$ than the minimum to $\theta = 1$. This is due to the fact that when θ is small each adsorbed molecule has z possible orientations, and when θ is large each adsorbed molecule has fewer possible ways of orientation due to the interference with the orientations with the neighbouring molecules.

---

* These equations are obtained by Peierls by using an aggregate with one central site, but they can also be obtained by considering aggregates with two central sites, as in the later part of § 2, or by considering two aggregates, one with one central site, and one with two, as in § 3.

Thus at large θ the probability of adsorption, apart from the effect of interactions, is smaller, and $d\log\lambda_{B_2}/d\theta$ is larger.

We describe roughly below calculations with a higher approximation for the case $z = 4$. We choose four sites forming a square as the central group of sites, each of them having two neighbours among the other sites of the group and two other neighbours outside the central group. These neighbours, eight in number, constitute the shell of neighbours surrounding the central group of sites. We first construct the g.p.f. for various configurations of the central group of sites exactly. For the two neighbours in the shell adjacent to a central site which is occupied by a B whose other end is among the central sites, we introduce a factor $(1+\lambda\eta\zeta)^2$ into the g.p.f. For the two neighbours in the shell adjacent to an unoccupied site, we introduce a factor $(1+\lambda\zeta)^2$. For the two neighbours adjacent to a central site which is occupied by a B whose other end is not among the other central sites, we introduce a factor $2\lambda\eta\zeta(1+\lambda\eta\zeta)$, corresponding to the fact that one of these neighbours must be occupied by the other end, and each possibility should be counted as a distinct state. It will be noted that we have put the same factor ζ into the g.p.f. for those B's in the shell whose other end is among the central sites, and for those whose other ends are outside the aggregate, and we have also excluded the possibility that a pair of neighbours in the shell be occupied by a molecule. Both approximations introduce errors into the results of the calculations, but we believe that the errors will not be serious in the present approximation.

We tabulate the various grand partition functions for future reference. Let $F(n_1, n_2)$ be the grand partition function for the configuration in which $n_1/2$ is the number of molecules B_2 lying entirely among the central sites, and n_2 the number of molecules with one end among the central sites, and the other end among the shell of neighbours. Then, writing $\epsilon = \lambda\zeta$, we have

$$\left.\begin{aligned}
F(4,0) &= 2\lambda^4\eta^4(1+\eta\epsilon)^8,\\
F(2,2) &= 4\lambda^4\eta^4(1+\eta\epsilon)^4\,[2\eta\epsilon(1+\eta\epsilon)]^2,\\
F(2,1) &= 8\lambda^3\eta^2(1+\eta\epsilon)^4\,(1+\epsilon)^2\,[2\eta\epsilon(1+\eta\epsilon)],\\
F(2,0) &= 4\lambda^2\eta(1+\eta\epsilon)^4\,(1+\epsilon)^4,\\
F(0,4) &= \lambda^4\eta^4[2\eta\epsilon(1+\eta\epsilon)]^4,\\
F(0,3) &= 4\lambda^3\eta^2[2\eta\epsilon(1+\eta\epsilon)]^3\,(1+\epsilon)^2,\\
F(0,2) &= \lambda^2[2+4\eta]\,(1+\epsilon)^4\,[2\eta\epsilon(1+\eta\epsilon)]^2,\\
F(0,1) &= 4\lambda(1+\epsilon)^6\,[2\eta\epsilon(1+\eta\epsilon)],\\
F(0,0) &= (1+\epsilon)^8.
\end{aligned}\right\} \qquad (44)$$

The equations concerned are

$$\theta = \frac{1}{4}\lambda\frac{\partial}{\partial\lambda}\log\sum_{n_1,n_2}F(n_1,n_2) = \frac{1}{8}\epsilon\frac{\partial}{\partial\epsilon}\log\sum_{n_1,n_2}F(n_1,n_2). \tag{45}$$

The result of solving (45) numerically is plotted below, with $\log\lambda^2\eta = \log\lambda_{B_2}$ as the ordinate, and θ as abscissa. For the adsorption of single atoms we need only replace the factor $2\eta\epsilon(1+\eta\epsilon)$ in (44) by $(1+\eta\epsilon)^2$, and in the summation in (45) we exclude all F except those whose n_1 is zero, everything else being the same. This has been done by G. P. Dube, who shows that the critical temperature is lower than that given by the first approximation,

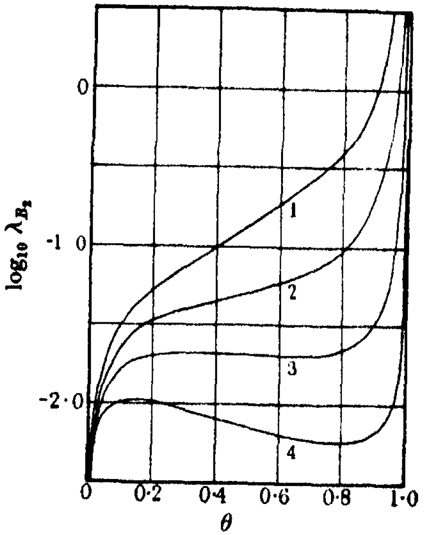


FIG. 3. Curves showing $\log_{10}\lambda_{B_2}$ against θ (obtained by second approximation) for attractive interactions. Curves: (1) $\eta = 1{\cdot}50$, (2) $\eta = 2{\cdot}04$, (3) $\eta = 2{\cdot}78$, (4) $\eta = 4{\cdot}00$.

i.e. $T_c = -(V/k)\{\log[z/(z-2)]^2\}^{-1}$. For our case, the critical temperature is given nearly enough by $\eta_c = 2{\cdot}78$, a value agreeing fairly well with the value 2·97 of the first approximation. Comparison of corresponding isotherms for the two approximations reveals that the isotherms behave generally in the same manner, with the tendency that here $\log\lambda_{B_2}$ is larger at small θ's, and smaller at θ's nearly unity.

We have not plotted the isotherms obtained by the present approximation which corresponds to repulsive interactions between the adsorbed molecules, as the curves are essentially monotonic, and so comparatively uninteresting. However, if the forces between the adsorbed molecules are repulsive, it is possible that the molecules arrange themselves regularly with the unoccupied sites, forming something like a superlattice, and thus the assumption employed throughout that the probability of adsorption is the same everywhere may not be valid (for superlattice formation of an adsorbed layer formed by single atoms, see Chang 1938). A special investigation into a particular type of superlattice shows that the temperature required for this to occur is comparatively low, so that this phenomena may be neglected for ordinary temperatures.

5. BRAGG AND WILLIAMS'S APPROXIMATION

Before applying Bragg and Williams's approximation to the present problem, we shall return first to our formula (22) for $g^*(\theta)$ obtained in § 2. In view of its importance and extreme simplicity of form, it is worth while to plot it and examine it roughly. At $\theta \fallingdotseq 0$, $g^*(\theta)$ varies as $-\frac{1}{2}\theta\log\theta$ and at $\theta \fallingdotseq 1$, $g^*(\theta)$ varies as $-(1-\theta)\log(1-\theta)$. It increases from zero at $\theta = 0$, with an infinite slope, reaches a maximum where $\theta(z-\theta)/(1-\theta)^2 = 1$, and decreases to a finite value $-\frac{1}{2}[(z-2)\log z-(z-1)\log(z-1)]$ at $\theta = 1$, also with an infinite slope.

For $z > 2$ the exact value of $g^*(\theta)$ has been found only for $\theta \fallingdotseq 0$, and $\theta \fallingdotseq 1$ (Fowler and Rushbrooke 1937). In our notation, the exact formula is

$$g(N,\theta) = \frac{\frac{1}{2}N\,!}{[\frac{1}{2}N\theta]\,!\,[\frac{1}{2}N(1-\theta)]\,!}\left(\frac{2z}{\sigma}\right)^{\frac{1}{2}N\theta}, \quad (\theta \fallingdotseq 0)$$

$$g(N,\theta) = \frac{N\,!}{[N\theta]\,!\,[N(1-\theta)]\,!}\left(\frac{z'''}{\sigma}\right)^{\frac{1}{2}N\theta}, \quad (\theta \fallingdotseq 1)$$

where σ is a symmetry number and is equal to 2 here, and z''' is a certain constant, depending upon the co-ordination number z only. Using Stirling's theorem we get

$$\left.\begin{aligned} g^*(\theta) &= -\tfrac{1}{2}\theta\log\theta-\tfrac{1}{2}(1-\theta)\log(1-\theta)+\tfrac{1}{2}\theta\log z, \quad (\theta \fallingdotseq 0)\\ g^*(\theta) &= -\theta\log\theta-(1-\theta)\log(1-\theta)+\tfrac{1}{2}\theta\log(z'''/2), \quad (\theta \fallingdotseq 1).\end{aligned}\right\} \qquad (46)$$

Here again, the most important part of the variation of $g^*(\theta)$ is $-\frac{1}{2}\theta\log\theta$ at $\theta \fallingdotseq 0$ and $-(1-\theta)\log(1-\theta)$ at $\theta \fallingdotseq 1$. The value of $g^*(\theta)$ at $\theta = 1$ is given by $\frac{1}{2}\log(z'''/2)$. Fowler and Rushbrooke have found that $z'''/2 = 1{\cdot}685$ for $z = 4$, and $2{\cdot}38$ for $z = 6$, and thus $g^*(\theta) = \log_e 10 \times 0{\cdot}1133$ for $z = 4$, and $\log_e 10 \times 0{\cdot}1883$ for $z = 6$. Formula (22) gives us

$$g(1) = -\tfrac{1}{2}[(z-2)\log z-(z-1)\log(z-1)],$$

which takes the value $\log_e 10 \times 0{\cdot}1136$ for $z = 4$, $\log_e 10 \times 0{\cdot}1911$ for $z = 6$, and $\log_e 10 \times 0{\cdot}2486$ for $z = 8$. In other words, the number of ways of arranging $\frac{1}{2}N$ double molecules upon N sites is given by Fowler and Rushbrooke as $(1\cdot685)^{\frac{1}{2}N}$ for $z = 4$, and $(2\cdot38)^{\frac{1}{2}N}$ for $z = 6$; and is given by formula (22) to be $(1\cdot688)^{\frac{1}{2}N}$ for $z = 4$, $(2\cdot41)^{\frac{1}{2}N}$ for $z = 6$, and $(3\cdot14)^{\frac{1}{2}N}$ for $z = 8$. The agreement is certainly satisfactory, in view of the simplicity of our formula, and the way in which it is obtained.

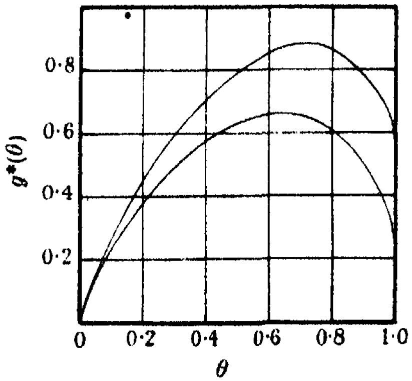


FIG. 4. Curves showing the number of ways, $e^{Ng^*(\theta)}$, of arranging $\frac{1}{2}N\theta$ double molecules upon N sites, with $g^*(\theta)$ plotted against θ: (1) upper curve for $z = 8$, (2) lower curve for $z = 4$.

A rough idea of the validity of (22) for intermediate values of θ can be obtained from comparing (22) and the exact values of $g^*(\theta)$ for the case $z = 2$, which can be obtained. In fact, for $z = 2$ we have

$$g(N, \theta) = (N - N\theta + \tfrac{1}{2}N\theta)!/(N - N\theta)!\,\tfrac{1}{2}N\theta!,$$

which is exact for all θ. Using Stirling's theorem, we get

$$g^*(\theta) = (1 - \tfrac{1}{2}\theta)\log(1 - \tfrac{1}{2}\theta) - (1 - \theta)\log(1 - \theta) - \tfrac{1}{2}\theta\log\tfrac{1}{2}\theta. \tag{47}$$

This is precisely (22) when $z = 2$. This agreement and the comparison in the above paragraph by no means establishes the validity of (22) for z different from 2, and all values of θ. At the same time, the approximation on which formula (22) is based seems to be equally reliable for all values of z and θ.

For clarity, we give the application of Bragg and Williams's approximation in detail. When there are $\frac{1}{2}N\theta$ molecules adsorbed, the partition function of the adsorbed phase is given by

$$\sum_X f(\tfrac{1}{2}N\theta, X)\, V_S^{\frac{1}{2}N\theta}\, e^{\frac{1}{2}N\theta\chi/kT}\, e^{-XV/kT}, \tag{48}$$

where X is the number of pairs of B-B nearest neighbours, belonging to different molecules, and $f(\frac{1}{2}N\theta, X)$ the corresponding number of configurations. According to the spirit of Bragg and Williams's approximations, we replace (48) by

$$\left.\begin{aligned} &\{\sum_X f(\tfrac{1}{2}N\theta, X)\}\, V_S^{\frac{1}{2}N\theta}\, e^{\frac{1}{2}N\theta\chi/kT}\, e^{-\bar{X}V/kT}, \\ \text{and assume that} \qquad &\bar{X} = N(z-1)\,\theta^2/2, \end{aligned}\right\} \tag{49}$$

since only $z-1$ neighbours are available to atoms of different molecules. To proceed further, we must know $\Sigma f(\tfrac{1}{2}N\theta, X)$, which is simply $g(N, \theta)$. The free energy of the adsorbed phase is

$$-kT\{\log g(N,\theta) + \tfrac{1}{2}N\theta(\chi/kT) + \tfrac{1}{2}N\theta \log V_S - N(z-1)\,\theta^2 V/2kT\}. \tag{50}$$

The free energy of the gaseous phase is

$$-N_g kT\left(\log \frac{(2\pi m_{B_2})^{\frac{3}{2}}\,(kT)^{\frac{3}{2}}\, V_g V}{N_g h^3} + 1\right). \tag{51}$$

Adding the free energies together, and taking the minimum with respect to variations of N_g or $\tfrac{1}{2}N\theta$, we get

$$\log \lambda_{B_2} + 2\frac{\partial}{\partial(N\theta)} \log g(N,\theta) - (z-1)\,2\theta\frac{V}{kT} = 0. \tag{52}$$

By using the limiting forms of Fowler and Rushbrooke, this reduces to

$$\left.\begin{aligned} &\log \lambda_{B_2} - (z-1)\,2\theta V/kT - \log[\theta/(1-\theta)\,z] = 0, \quad (\theta \fallingdotseq 0) \\ &\log \lambda_{B_2} - (z-1)\,2\theta V/kT - \log[2\theta^2/(1-\theta)^2 z'''] = 0. \quad (\theta \fallingdotseq 1) \end{aligned}\right\} \tag{53}$$

By using the formula (22), it reduces to

$$\log \lambda_{B_2} - (z-1)\,2\theta V/kT - \log[(z-\theta)\,\theta/z^2(1-\theta)^2] = 0. \tag{54}$$

It is easy to see that, as given by (54), the isotherms for positive and negative V and for various temperatures are not much different from those given by (32), (33). The critical temperature is, as before, determined by $d\lambda_{B_2}/d\theta = 0$ and $d^2\lambda_{B_2}/d\theta^2 = 0$, which gives $T_c = -1{\cdot}08\, V/k$ for $z = 4$, $-1{\cdot}76\, V/k$ for $z = 6$, $-2{\cdot}76\, V/k$ for $z = 8$, and $-(V/k)\,2z(3 + 2\sqrt{2})^{-1}$ for $z \to \infty$. Comparisons with the results in § 3 shows that the value of T_c is higher here, and the two values of T_c become the same as $z \to \infty$.

It is impossible to get the complete isotherm from (53). But it is easily seen that the parts of isotherms near $\theta \fallingdotseq 0$, and $\theta \fallingdotseq 1$, as given by (53) are similar to those given by (32), (33). We can also see roughly from (53) that the critical temperature is higher than $-\tfrac{1}{4}(z-1)(V/k)$. For if we take the isotherm to be represented by the first of (53) for the whole range of θ, we get, from using the equations $d\lambda_{B_2}/d\theta = d^2\lambda_{B_2}/d\theta^2 = 0$, $T_c = -\frac{1}{2}(z-1)(V/k)$, and if we take the isotherm to be represented by the second of (53) for the whole range of θ, we get $T_c = -\frac{1}{4}(z-1)(V/k)$. We have reasons to believe that the actual value is between the two, and is therefore greater than $-\frac{1}{4}(z-1)(V/k)$.

For completeness we may add that the adsorption isotherms with a partial dissociation among the adsorbed molecules behave intermediately between the isotherms with complete dissociation and those with no dissociations. For the particular case of no interactions between the adsorbed particles, equations (12) or (13) may remain with $(1-\theta)$ replaced by

$$(1-\theta)(1+\lambda e^{\frac{1}{2}W/kT}),$$

where W is the heat of dissociation. This really amounts to a distribution of those sites not occupied by double molecules into sites occupied by the dissociated atoms and unoccupied sites. The same result is obtained by enumerating different possible configurations of our typical aggregates, now with dissociated atoms, and proceeding exactly as in § 2. There will be no difficulty in so doing, as we need to introduce ζ only for those atoms among the neighbours which form molecules with atoms outside, all other atoms behaving to the outside as unoccupied sites.

Finally, we may add that the above calculations hold for the competitive adsorptions between double molecules B_2 and single atoms A, under the condition that the whole surface is covered by them. The equations are unchanged, but η will be replaced by $e^{-(V_{BB}+V_{AA}-2V_{AB})/kT}$, and λ by a quantity proportional to $(p_{B_2})^{\frac{1}{2}}/p_A$ and involving V_{BB} and V_{AA}. The details are uninteresting, and can be easily completed.

In concluding, the writer wishes to thank Professor R. H. Fowler for suggesting the problem and help given during the preparation of the paper, and G. S. Rushbrooke for discussions during the earlier developments.

[*Note added in proof.*] In a paper in preparation, the number of ways of arranging $N\theta\, B_2$ molecules upon N sites with given number of pairs of nearest neighbours formed by two B's of different molecules is found from (32) and (33). It is found to agree with the exact value for $z=2$. This means that (32) and (33) are exact for $z=2$ and justifies partly the procedure we adopt in § 3. The number of ways of arranging $N\theta$ single B atoms upon N sites with given number of nearest neighbours is also found there by making use of (43).

REFERENCES

Chang, T. S. 1938 *Proc. Camb. Phil. Soc.* **34**, 224.
Fowler, R. H. 1935 *Proc. Camb. Phil. Soc.* **31**, 260.
— 1936 *Proc. Camb. Phil. Soc.* **32**, 144.
— 1938 *Proc. Camb. Phil. Soc.* **34**, 382.
Fowler, R. H. and Rushbrooke, G. S. 1937 *Trans. Faraday Soc.* **33**, 1272.
Peierls, R. 1936 *Proc. Camb. Phil. Soc.* **32**, 477.
Roberts, J. K. 1937 *Proc. Roy. Soc.* A, **161**, 141.
Wang, J. S. 1937 *Proc. Roy. Soc.* A, **161**, 127.

Concerning the scattering of fast electrons and of cosmic-ray particles

BY E. J. WILLIAMS, *University College of Wales, Aberystwyth*

Formerly Leverhulme Foundation Fellow and Lecturer, Physics Department, University of Liverpool

(*Communicated by J. Chadwick, F.R.S.—Received* 25 *October* 1938)

When a particle, as the result of passing through a thickness of matter, is "scattered" through an angle, α, the scattering may be the result of the accumulated effect of a number of small deflexions produced by different atomic nuclei in the matter traversed, or it may be due to a single deflexion through an angle α produced by some one nucleus. The first type of scattering is spoken of as "multiple" or "plural", according as the number of contributing collisions is large or small. The second type is referred to as "single" scattering. Which process is mainly operative depends on the nature and velocity of the scattered particle, the matter traversed, and the scattering angle. Under the conditions of single scattering the experimental results may be readily compared with the theory of the scattering by a single nucleus or atom, and they have in that way a direct fundamental significance. The theory of multiple and plural scattering, in addition to the deduction of the single scattering formula, involves a statistical problem. An accurate treatment of the latter is in general not possible, and to the extent that this is so the corresponding experimental results are lacking in fundamental interest.

Much work has been done on the problem in connexion with the small-angle scattering of α-particles and also the scattering of β-particles. In these cases it has been shown that if thin foils are used as scatterers there is a region of single scattering even at small angles, though at *very* small angles the scattering is generally multiple or at least plural. The theory appropriate to the conditions of these experiments has been developed by Bothe (1929) and by Wentzel (1922), the calculations being based on a classical theory of collisions. These investigations, however, do not make possible an accurate quantitative interpretation of the observations (other than in the region of single scattering).

In the experiments on the scattering of cosmic-ray particles (Anderson 1933; Blackett and Wilson 1938), the scattering material is a metal plate a centimetre or so thick (placed in the path of the cosmic-ray particles in a Wilson chamber), while the particles have one electronic charge and a velocity very close to that of light. Under these conditions the scattering, as I have shown in an earlier note (Williams 1935*a*), is mainly multiple. This also applies to the recent experiments of Fowler and Oppenheimer (1938), and of Crane and Slawsky (1938), on the scattering of fast electrons (~1–20 million V energy) by thin metal foils. In view of the importance of a knowledge of the interaction of such high-energy particles with atomic nuclei it is very desirable to develop the statistical theory of the scattering sufficiently to enable an accurate interpretation of the results of the experiments to be made, which is not possible on the basis of existing treatments. The conditions of the experiments permit this to be done satisfactorily, and the purpose of this paper is, in the first place, to develop the theory in this respect, basing the statistical calculations on a quantum-mechanical theory of collisions. The main simplifying features of the experiments which make this possible are, (*a*) the average deflexion is chiefly contributed to by scattering through small angles, (*b*) the gaussian distribution representing the multiple scattering through small angles is nearly complete before the region of plural and single scattering is reached, (*c*) the scattering measured is the projection of the deflexion of the particle on a plane perpendicular to the line of sight and containing the initial direction. The arithmetic mean deflexion, which is the quantity most readily obtained in the experiments, may in fact be accurately evaluated irrespective of (*b*), though (*b*) is very helpful. The calculations are given in § 2 of this paper. In § 3 the results are used for a discussion of the experimental data. In this discussion reference is also made to the bearing of the cosmic-ray experiments on the existence of mesotrons ("heavy electrons").

An approximate treatment along the present lines was given in the note cited above. The formula then obtained for the arithmetic mean deflexion (which was the quantity considered by Anderson) could be in error by 10–20 %, but it was accurate enough for discussing the cosmic-ray results available at that time. The recent experiments of Blackett and Wilson have, however, greatly increased the scope and accuracy of the data, and the present calculations were undertaken chiefly in view of their work.

In the first part of this paper (§ 1) a brief discussion of the elements of the relativistic theory of scattering is given, and certain points are discussed in some detail. These include the effects of the shielding of the nucleus by the atomic electrons, and of the finite size of the nucleus. The first of these plays an essential part in multiple scattering, while under the conditions of the cosmic-ray experiments the second, rather unexpectedly, is almost equally important. It is found that observations on the scattering of cosmic-ray particles may also provide information about the non-electric interactions between these particles and the protons and neutrons in atomic nuclei. The data already available are in fact of some value in this respect.

The details of some of the calculations have been relegated to appendices.

1. RELATIVISTIC SCATTERING (EFFECT OF SHIELDING, ETC.)

(*a*) *General*

The scattering of fast electrons by a Coulombian field according to Dirac's relativistic theory has been calculated by Mott (1929). We require here the effect on the scattering of the shielding of the nucleus by the atomic electrons and of the finite size of the nucleus, and also the scattering arising from non-electrical interaction with the nucleus (both elastic and inelastic).

As we have already stated the angles of scattering involved are very small, especially in the cosmic-ray experiments where they are only of the order of a degree or so. We therefore only require the above modifications in the region of small angles, and this can be done fairly simply. Shielding in any case only concerns small angles for the high-energy particles under consideration.

In a field-free space a particle moving with velocity v parallel to a direction x may be represented by the infinite plane wave $Ae^{ix/\lambda}$, where $2\pi\lambda$ is the de Broglie wave-length. In the presence of a scattering field (electrostatic) in which the potential of the particle is $V(r)$, the solution will be of the form $\psi = Ae^{ix/\lambda} + \psi_1$, where ψ_1 represents the effect of the field. For low velocities ψ_1 is determined by the non-relativistic wave equation of Schrödinger, viz.

$$\left.\begin{aligned} \nabla^2\psi + (2M/\hbar^2)(E-V)\psi = 0, \quad E = \tfrac{1}{2}Mv^2, \\ \nabla^2\psi_1 + (2M/\hbar^2)E\psi_1 = (2M/\hbar^2)V(Ae^{ix/\lambda}+\psi_1). \end{aligned}\right\} \tag{1}$$

If the scattering field is sufficiently weak that the term in $V\psi_1$ may be neglected (Born's approximation) the solution for ψ_1, at large distances R from the scattering field, is given by

$$\psi_1(R) = \frac{A}{R}\frac{M}{2\pi\hbar^2}\int V(r)\, e^{i(x+|R-r|)/\bar{\lambda}}\, d\tau_r. \tag{2}$$

This means that the scattered wave at R may be obtained by supposing that every element of volume $d\tau_r$ of the field scatters waves of amplitude $(M/2\pi\hbar^2)\, V d\tau_r/R$, the phase of the secondary wave being the same as that of the primary wave at the point of scattering r. For a field that is spherically symmetrical the probability $I(\theta)$ of scattering through an angle θ, per unit solid angle (per incident particle per unit area), according to (2) is

$$\left.\begin{aligned} I(\theta) &= |\psi_1(R,\theta)|^2 R^2/A^2 = (2M/\mu\hbar^2)^2 f^2,\\ f &= \int V(r)\, r \sin\mu r\, dr, \quad \mu = 2\sin\tfrac{1}{2}\theta/\bar{\lambda}. \end{aligned}\right\} \tag{3}$$

For the scattering of a particle of charge e by a nucleus of charge Ze, (3) conforms to Rutherford's formula, viz.

$$I(\theta) = Z^2e^4/4M^2v^4\sin^4\tfrac{1}{2}\theta, \tag{4}$$

and if the field deviates from Coulombian the modified scattering may be obtained by evaluating (3) for the modified field. These are well known results in the non-relativistic theory of scattering.

The requirements of relativistic theory cannot in general be so simply expressed. However, for scattering through small angles, in which we are here primarily interested, it may be shown that the simple formula (3) may still be used if we insert the relativistic expression for the mass. Even for large angles, apart from spin effects, an expression of the type of (3) may be used.

Let us first consider the Schrödinger relativistic equation. For motion in an electrostatic field with a potential V, this equation is

$$\nabla^2\psi + \hbar^{-2}c^{-2}\{(E-V)^2 - M^2c^4\}\,\psi = 0, \tag{5}$$

$E = Mc^2(1-\beta^2)^{-\frac{1}{2}}$, $\beta = v/c$, $v =$ velocity in free space (i.e. where the scattering field $V = 0$). Written in terms of the de Broglie wave-length, $2\pi\bar{\lambda} = 2\pi\hbar(1-\beta^2)^{\frac{1}{2}}/Mv$, and substituting ξ for $(1-\beta^2)^{-\frac{1}{2}}$ it becomes

$$[\nabla^2 + \bar{\lambda}^{-2} - (2M\xi/\hbar^2)(V - V^2/2\xi Mc^2)]\,\psi = 0. \tag{5a}$$

In similar terms the non-relativistic equation is

$$[\nabla^2 + \bar{\lambda}^{-2} - (2M/\hbar^2)\, V]\,\psi = 0. \tag{1a}$$

The only difference in form is the appearance of $V-(V^2/2\xi Mc^2)$ for V. The new term $V^2/2\xi Mc^2$ arises from the variation of mass with velocity, and is the term responsible for the fine structure of spectral lines. In Dirac's relativistic theory for electrons there is, in addition to the V^2 term, a term representing the effect of spin. The second order wave equation in Dirac's theory is

$$[\nabla^2+\bar{\lambda}^{-2}-(2M\xi/\hbar^2)(V-V^2/2\xi Mc^2)-\overline{(i\partial V/\partial r/\hbar cr)\rho(\sigma,r)}]\psi=0, \qquad (6)$$

the last term being the extra one due to spin (cf. Mott 1929). At distances from the nucleus large compared with $\bar{\lambda}$ the new terms—V^2 and spin—are small compared with the V term, and it is this circumstance which simplifies the theory of scattering for small angles.

For a Coulombian field the ratio of the V^2 term to the V term is actually

$$(V/2\xi Mc^2)=(Ze^2/r2\xi Mc^2)=\beta(Ze^2/\hbar c)(\hbar/Mv\xi r)=(\beta Z/137)(\bar{\lambda}/r).$$

Thus even for the heaviest nuclei the ratio is always less than $\bar{\lambda}/r$. In the region of $r\gg\bar{\lambda}$ the V^2 term can therefore be neglected. This also applies to modified "Coulombian" fields provided the modification does not consist in an increase in the order of magnitude of V. As regards the spin term its ratio to the V term depends only on the distance r, and it follows from Mott's calculation that it is very small for $r\gg\bar{\lambda}$.

For a Coulombian field the scattering through an angle θ is due to the field at $r\sim\bar{\lambda}/\theta$.* Accordingly the scattering through small angles is due to the field at $r\gg\bar{\lambda}$. We have seen that at such r the spin term and the V^2 term in the relativistic equations are very small in comparison with the V term. The Coulombian scattering through small angles, and any modifications of it due to departures from a Coulombian field, may therefore be obtained by considering the effect of the V term only. Leaving out the other terms the relativistic equations reduce to

$$[\nabla^2+\bar{\lambda}^{-2}-(2M\xi/\hbar^2)V]\psi=0, \qquad (7)$$

which is identical in form with the non-relativistic equation (1*a*). The scattering, in Born's approximation, by analogy with the non-relativistic results (3), is therefore given by

$$I(\theta)=(2M\xi/\mu\hbar^2)^2f^2, \qquad (8a)$$

$$f=\int V(r)\,r\sin\mu r\,dr. \qquad (8b)$$

It is of interest to compare, for a Coulombian field, the relative effects of the different terms in the relativistic equations in terms of the "scattered" amplitudes to which they give rise. Denoting the amplitude of the incident wave by unity, the V term gives secondary waves of amplitude at unit distance from the nucleus (Rutherford formula) of

---

* This assumes $\alpha,=Z/137\beta,$ to be less than unity (cf. equ. (10)). For fast particles this is satisfied for all Z.

$$A_1 = (Ze^2/Mv^2\xi)\operatorname{cosec}^2 \tfrac{1}{2}\theta = A' \operatorname{cosec}^2 \tfrac{1}{2}\theta. \tag{9}$$

The amplitude due to the V^2 term may be obtained by substituting $(V^2/2Mc^2\xi)$ for V in (8). This gives, denoting $Ze^2/\hbar v$ by α,

$$A_2 = A'(\pi/2)\,\alpha\beta^2 \operatorname{cosec} \tfrac{1}{2}\theta(1 + \text{order}\,\alpha, \alpha^2, \ldots), \tag{11}$$

the terms of order α, α^2, ... in brackets denoting the uncertainty arising from the use of Born's approximation. As regards the spin term we find, working backwards from Mott's results for the scattering according to Dirac's theory and neglecting second order quantities, that it gives a scattered amplitude of

$$A_3 = -\tfrac{1}{2}A'\beta^2. \tag{12}$$

The respective amplitudes are thus in the ratio $1 : \frac{1}{2}\pi\alpha \sin \frac{1}{2}\theta : \frac{1}{2}\beta^2 \sin^2 \frac{1}{2}\theta$, which shows that for small angles the V^2 term and spin term have little effect, in conformity with what we have already pointed out.

The intensity of scattering is given by the square of the resultant amplitude. Thus for small angles

$$I(\theta) = A_1^2 = (Z^2e^4/4M^2\xi^2v^4 \sin^4 \tfrac{1}{2}\theta), \quad = R, \text{ say.} \tag{13}$$

The scattering according to Schrödinger's relativistic equation (all angles) is

$$I(\theta) = (A_1 + A_2)^2 = R\{1 + \pi\alpha\beta^2 \sin \tfrac{1}{2}\theta(1 + \text{order}\,\alpha, \alpha^2, \ldots)\}, \tag{14a}$$

and according to Dirac's theory, as given by Mott, it is

$$I(\theta) = R\{1 - \beta^2 \sin^2 \tfrac{1}{2}\theta + \pi\alpha\beta^2 \sin \tfrac{1}{2}\theta(1 - \sin^2 \tfrac{1}{2}\theta + \text{order}\,\alpha, \alpha^2, \ldots)\}. \tag{14b*}$$

(b) *Effect of shielding*

The potential field of an atom is little affected by the atomic electrons for distances, r, from the nucleus appreciably less than a, where

$$a = Z^{-\frac{1}{3}}a_0 = Z^{-\frac{1}{3}}(\hbar c/e^2)^2 (e^2/mc^2) = 137^2 Z^{-\frac{1}{3}}(e^2/mc^2). \tag{15}$$

a_0 is the radius of the first Bohr orbit of the hydrogen atom. For $r \gg a$ the nuclear field is completely neutralized by the electrons. An approximate expression sometimes used for the shielded field is

$$V = (Ze^2/r)\,e^{-r/a}. \tag{16}$$

---

* We see that apart from the term $(-\beta^2 \sin^2 \frac{1}{2}\theta)$ Mott's scattering formula follows in the main from the simple integration of (8b) with $(V - V^2/2\xi Mc^2)$ for V, where $V = Ze^2/r$.

Substituting this in the simplified formula (8) for the relativistic scattering gives $f = Ze^2\mu a^2(1+\mu^2a^2)^{-1}$ and hence

$$I(\theta) = R\{1+(\bar{\lambda}/2a\sin\tfrac{1}{2}\theta)^2\}^{-2}. \tag{17}$$

The shielding thus little affects the scattering for $\theta \gg \theta_{\min}$, where

$$\theta_{\min} \sim \bar{\lambda}/a, \tag{18}$$

and greatly reduces it for $\theta \ll \theta_{\min}$. This of course is an immediate consequence of the use of Born's approximation in which the scattering through an angle θ depends on the field at distances from the nucleus of the order of $\bar{\lambda}/\theta$.

The above result (equ. 17) involves the use of Born's approximation, which is justified in this connexion provided the quantity

$$\alpha = Ze^2/\hbar v \sim Z/137\,\beta \tag{10}$$

is appreciably less than unity.* For the heaviest elements α is not very small being about 0·6 for fast particles ($\beta \sim 1$), and appreciable errors might be expected to follow the use of Born's approximation under these conditions. This, however, is not so, because, whatever the value of α, the scattering by a pure Coulombian field, i.e. $a = \infty$ in (16), is exactly the same as that given by (17) when $a = \infty$. Accordingly the error of order α due to the use of Born's approximation concerns only the exact form of the falling off of the scattering in the region of $\theta = \lambda/a$.†

---

* This condition may be deduced as follows. The radius of the region around the nucleus responsible for the scattering of particles through angles greater than θ is of the order of $\bar{\lambda}/\theta$ in Born's approximation. Thus, if there is one incident particle per unit area, the number of incident particles traversing the scattering region is about $(\pi\bar{\lambda}^2/\theta^2)$. The number of particles scattered through angles greater than θ is less than the number scattered by an unshielded nucleus which, from the Rutherford formula, is $(4\pi Z^2e^4/M^2v^4\xi^2\theta^2)$. For Born's approximation to be valid the number of particles scattered by a given region (which measures ψ_1 in equ. 1) must be small compared with the number of particles incident on that region (which measures $Ae^{iz/\bar{\lambda}}$ in (1)), i.e. $(4\pi Z^2e^4/M^2v^4\xi^2\theta^2) \ll (\pi\bar{\lambda}^2/\theta^2)$, which gives $\alpha \ll 1$. This result also means that the condition underlying Born's approximation is essentially satisfied in the scattering by a pure Coulombian field if $\alpha \ll 1$. However a mathematically rigorous treatment of the problem requires more general methods than that represented by Born's approximation.

† If α were appreciably greater than unity the classical method of orbits would be valid for calculating the effect of shielding. Its application makes the scattering fall off at $\theta_{\min} \sim (Ze^2/\xi Mc^2\beta^2a) = \alpha(\bar{\lambda}/a)$. In the region of transition from the conditions of Born's approximation to those of classical orbits we have $\alpha \sim 1$, and the two methods give the same value of $\theta_{\min}$. This merging of the Born wave treatment with the classical method depends on the wave-length, $\bar{\lambda}$, being much less than the dimensions of the scattering region. If the reverse is the case as for instance in the collisions of slow neutrons with atomic nuclei the alternative to the simple wave treatment is of course that of Faxen and Holtsmark in which the incident wave is analysed into spherical harmonics.

For the calculation of the multiple scattering of cosmic-ray particles the approximate expression (16) for the atomic field is not quite accurate enough. The necessary degree of accuracy may be obtained by using the potential field given by the Thomas-Fermi distribution of atomic electrons. The non-relativistic scattering by such a field has been calculated by Bullard and Massey (1930). Since the relativistic equation we are using (equ. (7)) is of exactly the same form as the non-relativistic one (equ. (1*a*)) the results of Bullard and Massey may be used without further calculation. The scattering by a Thomas-Fermi atom may be expressed in the form

$$I(\theta) = R \times F(2a \sin \tfrac{1}{2}\theta/\lambda\!\!\!^{-}). \tag{19}$$

There is no analytical expression for the function F. Bullard and Massey have calculated it numerically for various values of the parameter, and the tabulated results given by them will be used when required in § 2.

The atomic electrons, besides shielding the nuclear field, give rise to an independent scattering in collisions in which they are ejected out of the atom or raised to a higher level. The intensity of such *inelastic* scattering is at the most only of the order of Z^{-1} of the elastic scattering given by (13), and it may therefore be neglected.*

(c) *Effect of finite size of nucleus*

The charge of an atomic nucleus is distributed amongst a number of centres, viz. the protons in the nucleus. If the scattering angle is sufficiently small that there is no interference between secondary waves from these different centres the scattering by the whole nucleus is the same as it would be if the charge were all concentrated at one point. However, if this is not so, i.e. if $(\lambda\!\!\!^{-}/\theta) \lesssim b$ where b represents nuclear dimensions, there is a phase difference between the secondary waves from the different centres within the nucleus and the scattering is thereby reduced. To calculate the scattering (elastic) under these conditions the distribution of charge may to a first approximation be assumed to be spherical and uniform. The electrical potential is then Ze^2/r outside the radius b of this sphere, and $(Ze^2/b)(\frac{3}{2}-r^2/2b^2)$ inside the sphere. A single expression which is roughly equivalent to this is

---

* The intensity of this electronic scattering follows readily from the theories of energy loss due to Bethe and Møller. For angles greater than $\lambda\!\!\!^{-}/a'$, where a' is the radius of the "orbit" of the atomic electron, the latter scatters as if it were free. Under these conditions it will scatter like a nucleus of unit charge up to angles for which it acquires a velocity comparable with that of the scattered particle. This limit is of the order of ξ^{-1} for the scattering of electrons, and of the order of m/M for the scattering of heavy particles of mass M, m being the electronic mass. There is no scattering through greater angles. For angles less than $\lambda\!\!\!^{-}/a'$ the scattering may be shown to be negligible compared with that by a free electron, i.e. by a nucleus of unit charge.

$$V(r) = (Ze^2/r)(1-e^{-2r/b}). \tag{20a}$$

Substituting this potential in (8) and integrating we have

$$I(\theta) = R\{1+(b\sin\tfrac{1}{2}\theta/\bar{\lambda})^2\}^{-2}. \tag{20b}$$ *

The deviation from a Coulombian field thus reduces the scattering by the factor within brackets. If $\bar{\lambda}$ is large compared with the nuclear dimensions represented by b, the factor is nearly unity for all θ.† For high-energy particles such as are met with in cosmic-rays $\bar{\lambda}$ is, however, much less than nuclear dimensions, and the correction is appreciable even for small angles of scattering.

It will be seen that whereas the correction for nuclear size reduces the scattering for θ of the order of and greater than $\bar{\lambda}/b$, the shielding reduces it for angles of the order of and less than $\bar{\lambda}/a$, where a represents atomic dimensions. Since $a \gg b$ there exists accordingly an extensive range of θ which obeys the Rutherford-like expression R given by (13). The corrections thus apply to regions which do not overlap. We shall, however, have occasion to consider cases where b has a different meaning and may be comparable with a. In that case we want the scattering by a field of the form $(Ze^2/r)\,e^{-r/a}(1-e^{-2r/b})$. The solution is

$$I(\theta) = R\{(1+\bar{\lambda}^2/4a^2\sin^2\tfrac{1}{2}\theta)^{-1}-(1+\bar{\lambda}^2(a^{-1}+2b^{-1})^2/4\sin^2\tfrac{1}{2}\theta)^{-1}\}^2. \tag{21}$$

Inside the nucleus there may be other than electric forces acting on the moving particle. Let us assume that these have an effective potential for the moving particle of the form

$$V(r) = Ve^{-r/b}, \quad = g(Ze^2/b)\,e^{-r/b} \text{ say,} \tag{22}$$

---

* It is of interest to observe that if the potential is taken equal to zero inside $r = b$, and to be Coulombian outside this radius the theoretical scattering is comparable with that for a pure Coulombian field for *all* angles, including the region of $\theta \gg \bar{\lambda}/b$. Actually such a sudden cut off reduces the scattering by a factor $\cos^2 x$, where $x = 2b\sin\frac{1}{2}\theta/\bar{\lambda} \sim b\theta/\bar{\lambda}$. This factor oscillates. It has an average value for $\theta \gg \bar{\lambda}/b$ of 1/2, while the reducing factor according to (20b) for such θ is $(\bar{\lambda}/\theta b)^4$. This result arises from the suddenness of the cut off, and reveals a certain arbitrariness in attributing scattering through θ to the region $r \sim \bar{\lambda}/\theta$. Analogous effects actually exist in optics such as for instance the illumination in the centre of the shadow cast by a circular disk. There is of course no very sharp cutting off (or rise) in the potential field of the nucleus and the results obtained using exponential expressions to represent the effect of the nucleus are quite valid in so far as this point is concerned.

† Under these conditions the correction is actually of a smaller order of magnitude than the effect of the V^2 term in (6), which we are already neglecting.

where g is positive or negative according as the electric and nuclear forces have the same or opposite sign. Adding this to the electric potential (equ. (20a)) we have a resultant $(Ze^2/r)(1-e^{-2r/b}+ge^{-r/b}\,r/b)$, giving

$$I(\theta) = R\{(1+\tfrac{1}{4}x^2)^{-1}+2gx^2(1+x^2)^{-2}\}^2, \tag{23}$$

where $x = 2b\sin\tfrac{1}{2}\theta/\bar{\lambda}$.

(*d*) *Inelastic nuclear scattering*

Nuclear collisions in which the scattering angle θ is much greater than $\bar{\lambda}/\delta$, where δ represents the dimensions of the cell occupied by each nuclear particle, result in the excitation or disintegration of the nucleus, the collisions being then "inelastic". In such collisions the nuclear particles scatter as if they were "free", and within the accuracy of the Born approximation this scattering is over and above the elastic scattering we have been considering, and does not take place at the expense of the latter. For $\theta \ll \bar{\lambda}/\delta$ the inelastic scattering is much less than if the nuclear particles were free. These points will be discussed more fully elsewhere.

Let us assume that the interaction between the incident particle and a nuclear particle is

$$V(r) = Ue^{-r/\sigma}+p(e^2/r)(1-e^{-2r/\sigma}), \tag{24a}$$

where $p = 0$ for a neutron, $p = \pm 1$ for a proton (according to the sign of the charge of the incident particle). Then for a nucleus of atomic number Z and mass number A the intensity of inelastic scattering through $\theta \gg \bar{\lambda}/\delta$, using the above result, is

$$\begin{aligned} I(\theta) = (R/Z)\{2uy^2(1+y^2)^{-2}+p(1+\tfrac{1}{4}y^2)^{-1}\}^2 \\ +\{(A/Z)-1\}(R/Z)\{2uy^2(1+y^2)^{-2}\}^2, \end{aligned} \tag{24b}$$

where $u = (U\sigma/e^2)$ and $y = 2\sigma\sin\tfrac{1}{2}\theta/\bar{\lambda}$. For $\theta \ll \bar{\lambda}/\delta$ the scattering is much less than is given by this expression.

The use of Born's approximation in the above calculations of the effect of the size of the nucleus, and the effect of nuclear forces, is justified only if the wave function in the scattering region does not differ appreciably from the undisturbed incident wave.* Assuming that the interaction energies do not exceed those between slow protons and neutrons, which are known approximately from the binding of light nuclei, we find that the condition is fairly well satisfied for the inelastic collisions, but is rather strained for the elastic scattering arising from the nuclear forces. The representation of the non-electrical interaction in terms of a simple potential (Wigner interaction) in the foregoing calculations is of course also not rigorous. The results are, however, sufficient for an approximate estimate of the effect of these forces on the multiple scattering of cosmic-ray particles and of fast electrons.

---

* In a recent treatment of inelastic collisions of fast particles with atomic nuclei Heisenberg (1937) assumes that Born's approximation is valid provided the energy of the incident particle is large compared with the binding energy of a nuclear particle. This of course is well satisfied when the incident particle is a cosmic-ray particle, but the condition given by Heisenberg is not the correct one for the validity of Born's approximation. The condition given above is much more stringent and is not necessarily satisfied even though the energy of the incident particle be infinitely large. In fact, for high-energy particles ($\beta \sim 1$) the energy is irrelevant.

2. MULTIPLE SCATTERING OF FAST ELECTRONS AND COSMIC-RAY PARTICLES

(*a*) *General*

In the last section we have given formulae for the scattering of fast particles through small angles. In the calculations the approximations $2\sin\frac{1}{2}\theta \sim \theta$ and $\cos\frac{1}{2}\theta \sim 1$ have not been used, because actually they would effect no simplification. In the problem of multiple scattering these approximations, however, make a difference and are essential to the present calculations. The errors they incur are negligible, being less than 1 % for the maximum angles involved, viz. $\theta \sim 10^\circ$. Introducing them into the Rutherford-like expression for the scattering (equ. (13)) it becomes

$$P(\theta)\,d\theta = 2k\,d\theta/\theta^3, \tag{25}$$

$$k = 4\pi NtZ^2e^4/M^2c^4\beta^4\xi^2. \tag{26}$$

t is the thickness of plate traversed, and N the number of nuclei per c.c. If the particle loses an appreciable fraction of its energy in traversing the plate so that β and ξ are not constant the effective value of $(1/\beta^4\xi^2)$ is its average value in the plate. In the cases that are considered in this paper, in which the loss of energy is small and takes place at a uniform rate the mean of the "incident" and "emergent" values of $(1/\beta^4\xi^2)$ may be taken, or even the reciprocal of the square of the mean values of $\beta^2\xi$.

A circumstance of the experiments is that it is the projection of the deflexion of the particles on the plane at right angles to the line of sight that is measured, this plane containing the initial direction. It is shown later (Appendix A) that if the distribution of θ is given by (25) the distribution of this projected deflexion, which we shall denote by ϕ, is

$$P(\phi)\,d\phi = kd\phi/\phi^3. \tag{27}$$

We now define an angle ϕ_1 such that in traversing the plate the particle experiences, on the average, *one* collision which deflects it through more than ϕ_1, i.e. $\int_{\phi_1}^{\pi} P(\phi)\,d\phi = 1$. $P(\phi)$, $= k/\phi^3$, decreases rapidly with ϕ so that the upper limit does not enter as long as ϕ_1 is small. Thus

$$\phi_1^2 = \tfrac{1}{2}k. \tag{28a}$$

Owing to the rapid variation of $P(\phi)$ with ϕ, collisions giving deflexions through angles appreciably less than ϕ_1 will take place in large numbers. These deflexions may be in the same or opposite direction, and it follows from the general theory of errors that the resultant deflexion, α_1, due to $\phi < \phi_1$, will conform very nearly to a gaussian distribution.* This may be written

$$P_1(\alpha_1)\,d\alpha_1 = (2/\pi\bar{\alpha}_1)\,e^{-\alpha_1^2/\pi\bar{\alpha}_1^2}\,d\alpha_1, \tag{29}$$

where $\bar{\alpha}_1$ is the arithmetic mean value of α_1. $\bar{\alpha}_1^2$ is $(2/\pi)$ times the mean square of α_1. The latter is the sum of the squares of the individual deflexions which build up α_1, and is accordingly equal to $\int_0^{\phi_1} \phi^2 P(\phi)\,d\phi$. Using the Rutherford-like expression (27) for $P(\phi)$ this gives

$$\bar{\alpha}_1^2 = \frac{2}{\pi}\int_0^{\phi_1} \phi^2 P(\phi)\,d\phi = \frac{2}{\pi}\int_0^{\phi_1} k\,d\phi/\phi = \frac{2k}{\pi}\,{}_0^{\phi_1}\Big[\log\phi\Big]. \tag{30a}$$

The integral becomes infinite if we take $\phi = 0$ as lower limit, and it is here that the shielding of the nuclear field by the atomic electrons comes in. This shielding makes the scattering negligible compared with the above Rutherford-like formula for angles appreciably less than $\phi_{\min} \sim \lambda\!\!\!^{-}/a$ (equ. (18)). This acts as an effective lower limit to the integral in (30*a*), and we have†

$$\bar{\alpha}_1^2 = (2k/\pi)\log(\phi_1/\phi_{\min}) = \phi_1^2(4/\pi)\log(\phi_1\,a/\lambda\!\!\!^{-}), \tag{30b}$$

$$= (4Z^2e^4Nt/M^2c^4\beta^4\xi^2)\log(Z^{\frac{4}{3}}Nth^2/2\pi m^2\beta^2c^2). \tag{30c}$$

For the scattering of cosmic-ray particles ($\beta = 1$) by 1 cm. of lead, ϕ_1 is about 300 times $\phi_{\min}$, making α_1 about three times ϕ_1. Thus the combined effect of the large number of deflexions through $\phi < \phi_1$ gives an average resultant much greater than ϕ_1. It is this which signifies the multiple nature of the scattering in this region. The gaussian distribution $P_1(\alpha_1)$ and the single scattering distribution $P(\phi)$ are plotted together in fig. 1. In this plot we take as unit of angle

---

* It is shown in Appendix B that this representation of the effect of $\phi < \phi_1$ is accurate within a few parts in a thousand.

† It is important to notice that since the shielding radius "a" of the atom is much less than the interatomic distances even in a solid, these distances and the arrangement of the atóms have a negligible effect on the scattering. The main crystal diffraction maxima take place at angles much smaller than α_1, where the atomic scattering is greatly reduced by shielding.

$$\delta = \sqrt{(k/\pi)} = 2\sqrt{(Nt)}\, Ze^2/Mc^2\beta^2\xi, \tag{31}$$

so that

$$\bar{\alpha}_1 = \sqrt{\{\log(\phi_1^2/\phi_{\min}^2)\}}\,(\delta), \tag{30d}$$

$$\phi_1 = \sqrt{(\pi/2)}\,(\delta). \tag{28b}$$

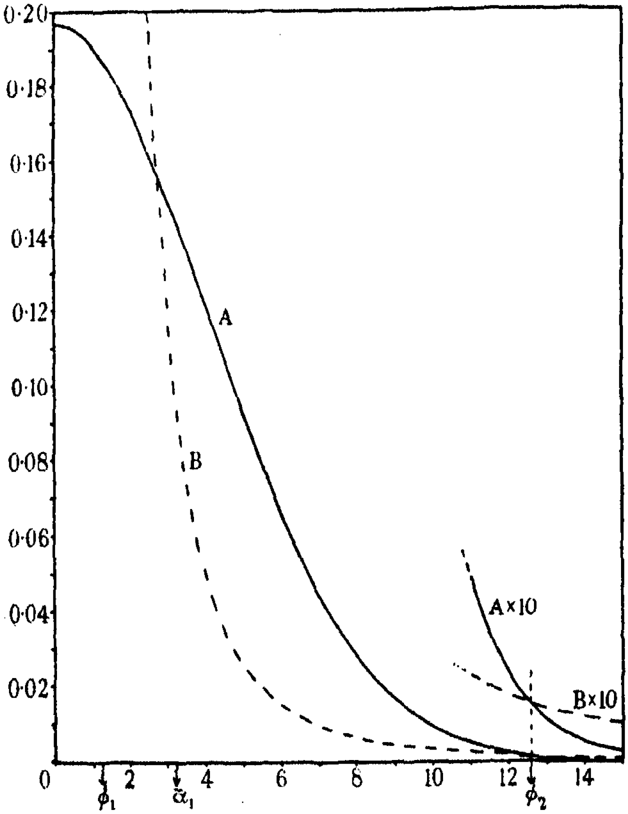


FIG. 1. *A* represents multiple scattering due to $\phi < \phi_1$ (equ. 29); $\bar{\alpha}_1 = 3{\cdot}23$. *B* represents single scattering formula $P(\phi) = \pi/\phi^3$. Abscissae represent angles in terms of δ as unit.

The scattering is single only for angles appreciably greater than ϕ_2 where the second intersection of the two curves takes place, and which in terms of δ as unit is given by

$$(2/\pi\bar{\alpha}_1)\, e^{-\phi_2^2/\pi\bar{\alpha}_1^2} = \pi/\phi_2^3. \tag{32}$$

The gaussian curve is practically complete before this angle is reached, which again shows that the average deflexion is almost entirely determined by multiple scattering.* This circumstance also greatly simplifies the analysis. The condition for this state of affairs, viz. that $\phi_2 > \bar{\alpha}_1$, and $\bar{\alpha}_1 > \phi_1$, is that the log term in (30*c*) be appreciably greater than unity. This requires

$$\phi_1 \gg \phi_{\min} \sim \lambda\!\!\!^{-}/a,$$

i.e.
$$M = (NtZ^{4/3}/\beta^2)(h^2/2\pi m^2c^2) \gg 1. \tag{32a}$$

In the case we have considered, viz. scattering by 1 cm. of lead ($\beta = 1$) M is of the order of 10^5. In the experiments of Fowler and Oppenheimer on the scattering of "artificial" high energy electrons the thickness of lead is only 0·013 cm. Even so M is still very large and the average deflexion is determined by multiple scattering. In the units chosen the curve for $P_1(\alpha_1)$ for 0·01 cm. is indeed much the same as that shown for 1 cm. of lead. In these units $P(\phi)$ is independent of the thickness.

The gaussian distribution $P_1(\alpha_1)$ represents the effect of nuclear deflexions through $\phi < \phi_1$. The occasional scattering through $\phi > \phi_1$ (one on the average) will contribute a little to the average deflexion, which will therefore be greater than $\bar{\alpha}_1$. This contribution was left out in my earlier calculations (1935*a*) and the expression (30*c*) for $\bar{\alpha}_1$ was given as the measure of the average deflexion. The order of magnitude of the additional effect of $\phi > \phi_1$ is $(\bar{\alpha}_1^2 + \phi_1^2)^{1/2} - \bar{\alpha}_1$, $\sim (\phi_1/\bar{\alpha}_1)^2 \times \bar{\alpha}_1$, so that the fractional error vanishes under the limiting condition of $\bar{\alpha}_1 \gg \phi_1$. In the cosmic-ray experiments however, $(\phi_1/\bar{\alpha}_1)^2 \sim 10\,\%$, and it is thus not inappreciable. This neglect of the effect of $\phi > \phi_1$ and the approximate treatment of the shielding factor, represent the two chief sources of error in the approximate value of the average deflexion given in the earlier treatment, and which is represented by (30*c*).

It is important to notice that the "mean square deflexion", $\overline{\alpha^2}$, as distinct from the linear mean ($\bar{\alpha}$), is in practice not free from an element of ambiguity if the nuclear field is Coulombian for the closest distances of approach to the centre of the nucleus concerned in the scattering. The mean square deflexion, arising from single scattering through $\phi < \phi''$, if ϕ'' is small, is

$$\overline{\alpha^2}'' = \int_0^{\phi''} \phi^2 \rho(\phi)\, d\phi. \tag{33}$$

---

* The average deflexion due to single scattering between ϕ_2 and ϕ' is $\int_{\phi_2}^{\phi'} (k/\phi^3)\,\phi\, d\phi$. The convergence of this integral as the upper limit, ϕ', is increased is important. It means that though the upper limit is actually π, only the region close to ϕ_2 is important. This is in contrast with the mean square deflexion represented by the integral $\int_{\phi_2}^{\phi'} (k/\phi^3)\,\phi^2\, d\phi$, which does not so converge. This is further discussed below.

It is of interest that, with the assumption of small angles, this expression is exact (whether the resultant distribution is gaussian or not).[*] This is so because a projected deflexion ϕ produced in a nuclear collision is as likely to be in one direction as in the opposite, so that, on the average, cross products have zero value when we square the algebraic sum of a number of deflexions ϕ.

Allowing for the effect of shielding in the approximate way already described, the integral in (33) becomes $\int_{\phi_{\min}}^{\phi''} \phi^2(k/\phi^3)d\phi = k\log(\phi''/\phi_{\min})$. This does not converge to a definite value as ϕ'' is increased, and it is this that introduces some ambiguity into $\overline{\alpha^2}$.[†] This lack of convergence is connected with the $1/\phi^3$ distribution of single scattering through small angles. If $P(\phi)$ decreased more rapidly with increasing ϕ the above integral would converge. It is here that the finite size of the nucleus plays an interesting part. The distribution k/ϕ^3 refers to a point charge. For a nucleus of radius b we have seen (equ. (20*b*)) that the scattering is very much less than is given by this expression for ϕ greater than about $\bar{\lambda}/b$. If therefore $\bar{\lambda}/b$ is *small*, as it is for cosmic-ray particles, it will act as an effective upper limit to the integral and we have

$$\overline{\alpha^2} = \int_{\phi_{\min}\sim\bar{\lambda}/a}^{\phi_{\max}\sim\bar{\lambda}/b} (k/\phi^3)\,\phi^2\,d\phi = k\log\left(\frac{a}{b}\right) = \pi\log(a/b)\,\delta^2. \tag{34}$$

Thus when $\bar{\lambda}/b \ll 1$, which for electrons is satisfied if their energy exceeds about 10^8 V, the mean square deflexion is determined by the ratio of the size of the atom (a) to the size of the nucleus (b). This state of affairs is represented diagrammatically in fig. 2, where the abscissae represent $\log\phi$ and the ordinates represent $\phi^3P(\phi)$. The area under the curve represents the mean square deflexion, and is seen to be limited by the "cut-offs" of the scattering effected by the shielding and by the finiteness of the nucleus.

---

[*] I am indebted to Dr H. Hulme for drawing my attention to this important point.

[†] In actual experiments an approximate effective upper limit for ϕ' could be made from a knowledge of the number of scattered particles observed, and the maximum observed deflexion, but the procedure would not be very straightforward. Of course there is no real ambiguity in $\overline{\alpha^2}$ since ϕ' cannot exceed π, but the use of a value of $\overline{\alpha^2}$ which is appreciably contributed to by single scattering through large angles ($\sim\frac{1}{2}\pi$) would be manifestly wrong if, as in the experiments concerned, the maximum deflexion—single or multiple—is only a few degrees. Such an expression would also be very difficult to estimate as the approximations $\sin\phi = \phi$, etc. could not be made, and it would be of use only when the number of observations is sufficient for all angles of scattering to be represented.

In the more exact calculations to follow, the various effects are considered in the following order: § 2(*b*) effect of shielding; (*c*) contribution of $\phi > \phi_1$ to the mean deflexion for point nucleus; (*d*) effect of size of nucleus and evaluation of mean square deflexion; (*e*) effect of non-electric interaction with the nucleus; (*f*) distribution of scattering.

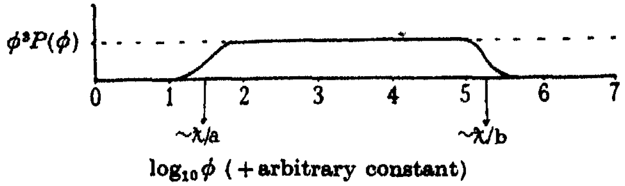


FIG. 2. Effect of shielding of nucleus and of its finite size on the single scattering.
- - - - Unshielded point nucleus. ——— Finite shielded nucleus.

(*b*) *Effect of shielding*

We require the value of $\int_0^{\phi_1} \phi^2 P(\phi)\, d\phi$. In the approximate treatment described in the last section $P(\phi)$ is taken equal to zero for $\phi < \lambda/a$, and equal to (k/ϕ^3) for $\phi > \lambda/a$. Actually of course the cut-off due to the shielding is spread over a range of angles in the neighbourhood of λ/a. The form of this cut-off is expressed by (19) where the function F can be obtained from tables given by Bullard and Massey. Since their calculations refer to the actual angle of scattering θ, and not its projection ϕ, we first evaluate $\int_0^{\phi_1} \theta^2 P(\theta)\, d\theta$ using their values of $F(2a \sin \frac{1}{2}\theta/\lambda)$, and then convert back into ϕ. The details of the calculation are described in Appendix A. The result obtained corresponds to a value of the effective lower limit to ϕ of

$$\phi_{\min} = 1{\cdot}75(\lambda/a), \tag{35}$$

giving
$$\overline{\alpha}_1^2 = (2/\pi)\int_{\phi_{\min}}^{\phi_1} \phi^2(k/\phi^3)\, d\phi = 2\log(\phi_1 a/1{\cdot}75\lambda)\,\delta^2 \tag{36}$$
$$= \log(Z^{4/3}Nt\hbar^2/1{\cdot}75^2 2\pi m^2\beta^2c^2)\delta^2.$$

The difference between this and the earlier approximate estimate (equ. (30*b*)) is represented by the factor 1·75, which decreases $\overline{\alpha}_1$ by about 4 %.

(*c*) *Allowance for scattering through* $\phi > \phi_1$

This is considered in detail in Appendix B. The final formula obtained for the arithmetic mean deflexion with the above value of $\phi_{\min}$ is

$$\overline{\alpha} = \{3{\cdot}69 + 0{\cdot}28\log_{10}(Z^{\frac{4}{3}}\rho t/W\beta^2)\}\,\delta. \tag{37}$$

ρ = density of scattering plate, W = atomic weight of material of plate. The earlier approximate formula (equ. (30)), which may be expressed in the form

$$\bar{\alpha}_1 = \{3\cdot04 + 0\cdot34 \log_{10}(Z^{\frac{4}{3}}\rho t/W\beta^2)\}\,\delta \qquad (30e)$$

is about 20 % less. The computational error in (37) (see Appendix C) does not exceed about 1 % provided $\bar{\alpha}$ is of the order of or less than about 10°, and provided that $(Z^{\frac{4}{3}}\rho t/W\beta^2)$, which is about 20 for 1 cm. of lead and $\beta = 1$, lies between about 0·1 and 10^4 (which covers the entire experimental region).*

(*d*) *Effect of size of nucleus, and the root-mean-square deflexion*

Equation (37) is based on the assumption of a point nucleus, i.e. a Coulombian field down to $r = 0$. We have seen that the finite size of the nucleus (radius b) makes the scattering through angles of the order of and greater than λ/b very small compared with the pure Coulombian scattering for such angles. If λ/b is appreciably greater than the angle ϕ_2 (fig. 1), the average angle of scattering is however little affected and is practically the same as that for a point nucleus given by (37). The reason for this is that the contribution, to the average deflexion, of single scattering through angles appreciably greater than ϕ_2 is negligible. In fact the whole region above ϕ_2 even for the unmodified Coulombian field contributes only about 4 % (see equ. (44)). However, if λ/b is appreciably less than ϕ_2 there is a sensible difference between the average deflexion for a point nucleus and that for a nucleus of radius b. Taking b for lead $\sim 1\cdot0\times10^{-12}$ cm., we find that these conditions obtain in the cosmic-ray experiments (see subsection (*f*) where $\alpha_{ms} \sim \phi_2$ is compared with λ/b). Under these conditions the arithmetic mean deflexion is given by

$$\bar{\alpha}_b^2 = (2/\pi)\int P(\phi)\,\phi^2\,d\phi\dagger = (1/\pi)\int P(\theta)\,\theta^2\,d\theta.\ddagger$$

We can use the formula for $P(\theta)$ for a Coulombian field, viz. $2k/\theta^3$ (equ. (25)) provided we introduce the limits ϕ_{min}, $\sim\lambda/a$, and ϕ_{max}, $\sim\lambda/b$, these limits representing the effects of shielding and nuclear size respectively. Thus $\bar{\alpha}_b^2 = (2k/\pi)\log(\theta_{\text{max}}/\theta_{\text{min}}) = (2k/\pi)\log(\sim a/b)$, and is determined by the ratio of the size of the atom to the size of the nucleus. The accurate treatment of shielding in Appendix A (equ. (i)) gives

---

* Cf. footnote *, p. 550.

† See end of Appendix B.

‡ ϕ^2 is on the average equal to $\frac{1}{2}\theta^2$, θ being the actual angle of scattering, ϕ its projection.

$$\theta_{\min} = 2\cdot 10(\lambda\!\!\!^{-}/a). \tag{35b}$$

Using the formula (20b) for the scattering by a nucleus of radius b gives

$$\theta_{\max} = 1\cdot 20(\lambda\!\!\!^{-}/b). \tag{38}$$

Hence $\overline{\alpha_b^2} = (2/\pi)\,k\log(0\cdot 6a/b)$. On substituting for a (equ. (15)) and taking

$$b = 0\cdot 7Z^{\frac{1}{3}}(e^2/mc^2), \tag{39}$$

$$\overline{\alpha}_b = (19\cdot 5 - 3\cdot 1\log_{10} Z)^{\frac{1}{2}}\,\delta. \tag{40}$$

The condition $\lambda\!\!\!^{-}/b \ll \phi_2$ underlying this result is sufficiently satisfied to make it accurate within about 2 % if $M' > 0\cdot 3\beta M'_0$, where $M' = Z^{\frac{4}{3}}\sqrt{(Nt)}$, and M'_0 is the value of M' for 1 cm. of lead. An error of about 1 or 2 % may also arise from the approximate nature of the expression (20a) used to represent the electrical field of the nucleus.

It has been shown in § 2a that the reduction in the scattering for angles greater than $\lambda\!\!\!^{-}/b$, effected by the finite size of the nucleus, serves to give to the mean square deflexion, $\overline{\alpha_b^2}$, an unambiguous value provided $\lambda\!\!\!^{-}/b$ is small (but not necessarily less than ϕ_2). This value, as shown in § 2a, is accurately given by $\int \phi^2 P(\phi)\,d\phi$. This is $\frac{1}{2}\pi$ times the above estimate of $\overline{\alpha_b^2}$, so that from (40)

$$\sqrt{\overline{\alpha_b^2}} = (\tfrac{1}{2}\pi)^{\frac{1}{2}}(19\cdot 5 - 3\cdot 1\log_{10} Z)^{\frac{1}{2}}\,\delta. \tag{41}$$

(e) *Effect of non-electric forces*

When the moving particle passes through, or very close to an atomic nucleus, it may be deflected by other than electric forces. This effect will concern even small angle scattering when $\lambda\!\!\!^{-}$ is appreciably less than nuclear dimensions, which as we have seen obtains in the cosmic-ray experiments. The magnitude of the non-electric forces is not known from experiment, except when the incident particle is a neutron or proton. In the observations on the scattering of cosmic-rays most of the particles of medium energy are probably of the new type ("mesotrons"), with electrons predominating at lower energies (~ 10^8 V), and possibly protons amongst the higher energy ones ($\gtrsim 5 \times 10^9$ V). If the non-electrical interaction between these particles and the nuclear particles were comparable with that found experimentally between slow neutrons and protons (deduced from the binding energy of light nuclei, etc.), the effect on the scattering under the conditions of the cosmic-ray experiments would be appreciable.

The effect of *inelastic* scattering is, in this connexion, greater than that of the elastic scattering, especially if we measure the scattering in terms of the mean square deflexion. Denoting the intensity of the inelastic scattering by $P_{\mathrm{in}}(\theta)$ its contribution, $\overline{\alpha^2_{\mathrm{in}}}$, to the mean square projected deflexion is $\frac{1}{2}\int \theta^2 P_{\mathrm{in}}(\theta)\, d\theta$,* which using the equation (24*b*) for $P_{\mathrm{in}}(\theta)$ for $\theta > \lambda\!\!\!^{-}/\delta'$† (and neglecting the electrical interaction with protons, i.e. $p = 0$) gives

$$\overline{\alpha^2_{\mathrm{in}}} = \frac{\pi}{3}\left(\frac{j}{Z}\right)^2 \frac{1+3z^2}{(1+z^2)^3}\,\delta^2, \tag{42}$$

where $j = A^{\frac{1}{3}}(U\sigma/e^2)$ and $z = \sigma/\delta'$.

A similar calculation gives the effect of the elastic scattering due to the nuclear forces (equ. (23)) as

$$\overline{\alpha^2_{\mathrm{el}}} = \pi\,(\tfrac{1}{3}g^2 + 1{\cdot}4g)\,\delta^2 \tag{43‡}$$

where $g = Vb/Ze^2$.

(*f*) *Distribution of scattering*

The curve A in fig. 1 represents the distribution of the resultant scattering due to single scattering through $\phi < \phi_1$, on the assumption that it is gaussian (equ. (29)), with an average deflexion given by (36). It is shown in Appendix C that this gaussian representation is correct within a few parts in a 1000. The scattering through $\phi > \phi_1$ but less than ϕ_2, which marks the intersection of A and the single scattering distribution B (equ. (32)), causes a certain broadening of the distribution. Since this broadening is small ($\sim 20\,\%$), and since in the neighbourhood of ϕ_2 the curve A is very steep, and much steeper than B, the broadened distribution is still very nearly gaussian. The average deflexion, α_m, corresponding to it is the average deflexion $\bar{\alpha}$ due to all ϕ (equ. (37)), *minus* the contribution of $\phi > \phi_2$. Since ϕ_2 is much greater than $\bar{\alpha}$ this contribution is (in units δ)

$$\alpha_s = \int_{\phi_2} \phi(\pi/\phi^3)\, d\phi = \pi/\phi_2, \tag{44}$$

---

* The factor 1/2 allows for the fact that it is the mean square of the *projected* deflexion that we require.

† δ in (24) is written δ' here to distinguish it from the unit of angle δ.

‡ In allowing for the elastic scattering due to the nuclear forces it must be remembered that the amplitude of this scattering has to be added to that due to the electric forces, and *then* squared to give the intensity of scattering. The term $1{\cdot}4g$ in (43) arises from the cross product in the square. Its average value for positive and negative particles is zero if we assume the sign of the nuclear forces to be independent of the charge.

and therefore
$$\alpha_m = \bar{\alpha} - \pi/\phi_2. \tag{45}$$ [*]

The resultant distribution $P_r(\alpha)$ is now very nearly the new gaussian distribution, with mean deflexion $= \alpha_m$, i.e.

$$P_m(\alpha) = (2/\pi\alpha_m)\, e^{-\alpha^2/\pi\alpha_m^2} \tag{46}$$

up to $\alpha = \phi_2$, and the sum of $P_m(\alpha)$ and the single scattering distribution $P_s(\alpha) = \pi/\alpha^3$ beyond ϕ_2, with of course a "smoothing" in the region of ϕ_2[†] (see fig. 3). The exact form of the curve near ϕ_2 is, however, of little practical importance as its height in this region is very small, about 1/100th of the maximum at $\alpha = 0$. The gaussian component P_m intersects the single scattering distribution P_s at $\alpha = \alpha_{ms}$ say, given by

$$(2/\pi\alpha_m)\, e^{-\alpha_{ms}^2/\pi\alpha_m^2} = \pi/\alpha_{ms}^3. \tag{47}$$

α_{ms} is about 20 % greater than ϕ_2, and for 1 cm. of lead its value in terms of δ is about 15. The gaussian curve is very steep in this region and consequently there is a fairly sharp transition from multiple to single scattering in the neighbourhood of α_{ms}. "Plural" scattering is thus limited to a very small region of angles.

The effect of the finite size of the nucleus on the distribution depends on the relative values of α_{ms} and the angle λ/b, where b is the radius of the nucleus. We have seen that this effect greatly reduces the single scattering for ϕ greater than about λ/b (cf. (20b)). Thus if λ/b is appreciably less than α_{ms} the single scattering "tail" beyond α_{ms} is cut out almost completely, and the resultant distribution is practically gaussian with an average angle of scattering given by (40). If on the other hand λ/b is greater than α_{ms} the effect is simply a modification of the single scattering tail (curve P_s) to conform to (20b).

For fast particles ($\beta \sim 1$) the ratio of λ/b to α_{ms} is independent of the energy of the incident particle, and is approximately equal to $4{\cdot}0/Zb\sqrt{(Nt)}$. For lead, taking $b = 8 \times 10^{-13}$ cm., this is about $(0{\cdot}3/\sqrt{(t)})$. In most of the cosmic-ray experiments $t \sim 1$ cm., so that $\lambda/b \sim 0{\cdot}3\alpha_{ms}$. We thus have the first of the

---

[*] It happens that α_m is given with considerable accuracy (error < 1 %) by $\alpha_m^2 = 2\log(\phi_2/\phi_{\min})$, which is similar to the expression (30d) for α_1^2, except that ϕ_2 replaces ϕ_1. The mean deflexion $\bar{\alpha}$ due to all ϕ may thus be given an alternative expression to (37), viz.

$$\bar{\alpha} = \{2\log(\phi_2/\phi_{\min})\}^{\frac{1}{2}} + \pi/\phi_2. \tag{37a}$$

This indeed may be a more accurate expression when the conditions are such that M (equ. (32a)) differs in several orders of magnitude from its value for 1 cm. of lead, $\beta = 1$.

[†] The total area under the composite curve thus formed is slightly greater than unity ($\sim 1{\cdot}008$). The difference from unity is however so small that it is unnecessary to introduce any correcting factors.

conditions mentioned above, viz. $\lambda\!\!\!^{-}/b < \alpha_{ms}$. The single scattering tail is cut out. For $t \sim 0{\cdot}01$ cm., as in the recent experiments of Fowler and Oppenheimer $\lambda\!\!\!^{-}/b \sim 3\alpha_{ms}$, and we have the second condition, the finiteness of the nucleus only modifying the single scattering tail at angles much larger than α_{ms}. In their experiments the energy of the electrons under observation is however such that α_{ms} is about 30°, and $\lambda\!\!\!^{-}/b$ is in a region of large angles not investigated in the experiments.

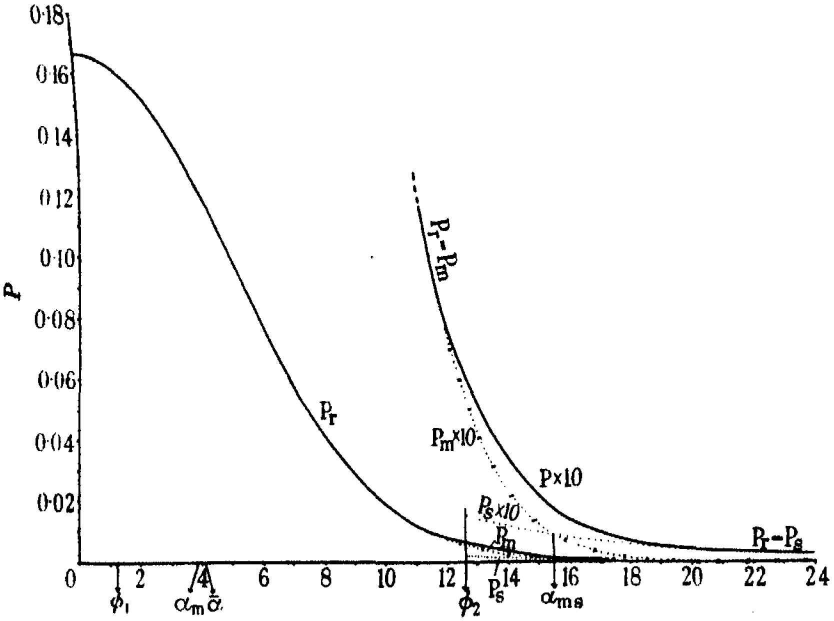


FIG. 3. Theoretical distribution of scattering.

× · · · × · · · × $P_m(\alpha) = (2/\pi\alpha_m)e^{-\alpha^2/\pi\alpha^2_m}$, $\alpha_m = 3{\cdot}82$ (multiple)
- - - - - - $P_s(\alpha) = \pi/\alpha^3$ (single)
——— $P_r(\alpha) =$ Resultant distribution.

Finally there is the effect of non-electric forces inside the nucleus. If these are of the same sign as the electric force, they tend to compensate for the absence of a Coulombian electric field inside the nucleus, and thereby restore to some extent the scattering through angles of the order of $\lambda\!\!\!^{-}/b$ (equ. (23)). This however does not apply to $\phi \gg \lambda\!\!\!^{-}/b$. Even if we assume the nuclear forces to be as strong as those between slow protons and neutrons, which is unlikely to be the case for cosmic-ray particles, the restoration of the elastic scattering by the nuclear forces is insignificant for ϕ greater than about twice $\lambda\!\!\!^{-}/b$.

Inelastic collisions, as already indicated in section (*e*), are more likely to give evidence of non-electrical interaction with the nuclear particles if this interaction is of short "range". Assuming the interaction to be of the form $Ue^{-r/\sigma}$ (equ. (24*a*)) the intensity of inelastic scattering is (equ. (24*b*) with $p=0$)

$$P_{\text{in}}(\phi)=\frac{k}{\phi^3}\frac{5}{2}\left(\frac{j}{Z}\right)^2\frac{z^3}{(1+z^2)^{\frac{5}{2}}};\quad z=\phi\sigma/\lambda\!\!\!^{-},\ j=U\sigma A^{\frac{1}{3}}/e^2,\qquad (48)^*$$

except for small $\phi(\ll\lambda\!\!\!^{-}/\sigma)$, where the scattering is much less than this expression gives. The expression is a maximum at $z=0$, and bears its maximum ratio to the Coulombian scattering k/ϕ^3 at $z=\sqrt{3/2}$, i.e. ϕ, $=\phi_{\text{in}}^m$, $=(\sqrt{3/2})(\lambda\!\!\!^{-}/\sigma)$, $\sim(3\cdot10^{-13}/\sigma\sqrt{t})\,\alpha_{ms}$ for the scattering by lead.

In the discussions to follow we shall use the interaction between slow neutrons and protons as a kind of yardstick for non-electrical interaction. This neutron-proton interaction is known approximately from the binding energy of light nuclei, and the scattering of slow neutrons by protons. Assuming it to be of the above form, viz. $U'e^{-r/\sigma'}$ (and averaging over spin orientation) the data give

$$U'\sigma'^2\sim12(e^4/mc^2),\quad \sigma'\sim1\cdot2\times10^{-13}.\qquad (49)$$

The value of σ' is less certain than that of $U'\sigma'^2$. With these values we find that for the scattering by a 1 cm. lead plate the expression (48) bears its maximum ratio to Coulombian scattering at ϕ, $=\phi_{\text{in}}^m$, $\sim2\alpha_{ms}$, the ratio at this angle being about 6.

It is of considerable interest that for short range forces ($\sigma\sim10^{-13}$ cm.) ϕ_{in}^m occurs well beyond α_{ms}, for at such angles the scattering due to electric forces (when the structure of the nucleus is taken into account) is very small. Actually the probability of scattering, both elastic and inelastic, through $\phi>1\cdot3\alpha_{ms}$ due to electrical interaction alone is only about 0·0001. This absence of "background" due to electrical effects makes the region particularly suitable for the investigation of the non-electric interaction between cosmic-ray particles and nuclear particles. The absolute probability of scattering into this region due to such interaction is unfortunately rather low ($\sim0\cdot02$ assuming the interaction represented by (49)). The possible methods of attack on this problem are however limited and the investigation of the scattering in the above region may prove valuable.

---

* (24*b*) refers to the actual scattering angle θ. (48) is the corresponding distribution of ϕ, assuming the angles to be small. The average projected deflexion due to this inelastic scattering (by itself) is

$$\overline{\alpha_{\text{in}}}=(\pi/2)(j/z)^2(\sigma/\lambda\!\!\!^{-})(1+z_1^2)^{-\frac{3}{2}}\delta^2,\qquad (48a)$$

where $z=z_1(\sim0\cdot7)$ is taken as an effective lower limit to the applicability of (48).

The relation of non-electrical scattering assuming the above proton-neutron interaction to the electrical scattering is shown diagrammatically in fig. 4. The ordinates represent the intensity of scattering in terms of that given by the Rutherford-like expression, R (equ. (13)), as unit. (The area under the curves measures the contribution to the mean square deflection.)

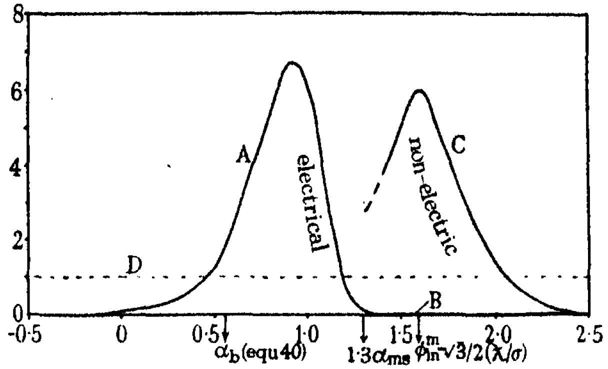


FIG. 4. Showing scattering by 1 cm. lead plate due to "non-electrical" and electrical forces respectively.

A, elastic scattering due to electrical interaction (multiple); B, inelastic scattering due to electrical interaction (height of curve $\ngtr Z^{-1} = 1/82$) (single); C, inelastic scattering assuming proton-neutron like non-electrical interaction (equ. 49) (single); D, single scattering formula for unshielded point nucleus. Abscissae: log (deflexion in units δ); ordinates: scattering probability in terms of scattering by unshielded point nucleus as unit (Rutherford-like formula).

3. COMPARISON WITH EXPERIMENT

(a) *Cosmic-ray region*

In the experiments on the scattering of cosmic-ray particles the radius of curvature (ρ) of the tracks of the particles in a magnetic field (H) is measured. $H \times \rho = M\beta c^2\xi/e$, where M is the mass of the particle, βc its velocity, and $\xi = (1-\beta^2)^{-\frac{1}{2}}$. Instead of "$H\rho$" itself the quantity usually recorded is the energy, E, in volts, calculated from the value of $H\rho$ assuming $\xi \gg 1$. $E = 300H\rho$. In the foregoing calculations we have used $\delta = (2Ze^2\sqrt{(Nt)}/M\beta^2c^2\xi)$ (equ. (31)) as a unit of angle. In terms of E

$$\delta = (600Ze\sqrt{(Nt)}/\beta E) = \delta_0/\beta. \qquad (31a)$$

We see that for a given value of E, δ is independent of M except in so far as M enters into the dependence of β on E. This also applies to the factors by which δ has to be multiplied to give the mean or mean square deflexion (equ. (37), (40–43)). This is an important result because cosmic-ray particles are of different kinds, with different masses. Their energy, as may be judged from their specific ionization, is however, in general, sufficiently high that β is very close to unity. In that case, for a particle with a given "$H\rho$" the mass need not be known at all for calculating the scattering.

The theoretical arithmetic mean deflexion, $\bar{\alpha}$, for a point nucleus (equ. (37)) is equal to $\{3{\cdot}69 + 0{\cdot}28 \log(Z^{\frac{4}{3}}\rho t/W\beta^2)\}\delta$, which for the scattering by 1 cm. of lead $= (4{\cdot}05/\beta)\,\delta_0$. The dotted horizontal line A in fig. 5 represents this for $\beta = 1$. For electrons the deviation from this line is too small to be shown. Curve B represents the scattering of particles with a mass about one-tenth that of a proton (mesotrons).* Even this deviates very little from the horizontal line A in the region above 2×10^8.

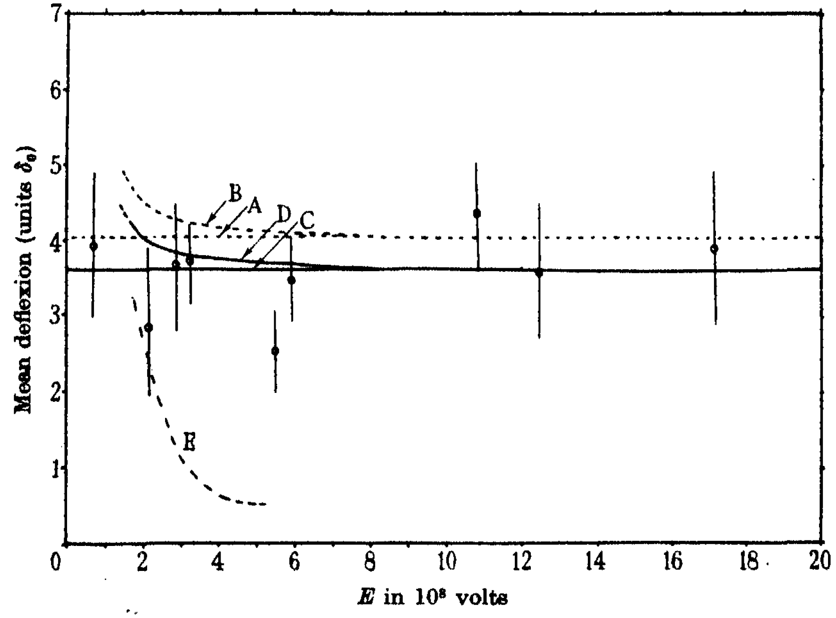


FIG. 5. Scattering of cosmic-ray particles. ⏀-experimental results of Blackett and Wilson. Curves represent theoretical values; A for point nucleus and electrons, B for point nucleus and mesotrons, C and D for electrons and mesotrons respectively allowing for finite size of nucleus, E for electrons assuming low energy loss to be due to breakdown of theory.

The experimental results of Blackett and Wilson are represented in the figure by ϕ, the vertical line indicating the probable experimental error.* It will be seen that there is a fairly satisfactory agreement with the theoretical values, as in fact Blackett and Wilson have already shown, using the approximate formula (30*c*) for $\bar{\alpha}$ (which does not differ much from the more accurate one used here). We have seen in § 1 that by virtue of the scattering being through small angles the doubtful relativistic elements connected with "spin" and the "V^2" term (equ. (5*a*) and (6)) do not enter into the calculations. The general agreement with experiment is therefore to be expected.

---

* The curves are stopped in the region of energy where they lose in the plate a large fraction of their energy by ionization.

We shall here digress a little to explain the significance of the curve (*E*). Before the suggestion of the existence of mesotrons it was thought that cosmic-ray particles were mainly electrons, and the low energy loss of the particles with $E > 2 \times 10^8$ V, as compared with the theoretical value for electrons, was attributed to a breakdown of the theory of the energy loss.† It was pointed out by the writer at the time (1935*a*) that it would be difficult to conceive a type of breakdown that would not also reduce the theoretical scattering, which even the early measurements of Anderson showed to be in approximate agreement with theory. Curve *E* in fig. 5 represents the expected scattering on a "breakdown" hypothesis. It is calculated for a nuclear field modified so as to give an energy loss for electrons equal to that observed for the cosmic-ray particles.‡ It is seen that in the region of 4×10^8 V this gives a scattering only about 10 % of the observed. This result, as I have recently shown elsewhere (1938*a*), constitutes very strong evidence against the breakdown hypothesis, and therefore in favour of the hypothesis that the low energy loss is due to the particles being of a new type—the mesotron—with a mass much greater than that of electrons. The latter hypothesis makes the energy loss much less than the theoretical value for electrons (as required), but leaves the scattering unchanged because as we have seen the scattering of fast particles of a given "$H\rho$"

---

* Some of the points refer to the scattering by copper (2 cm. plate). These have been reduced to the conditions of 1 cm. of lead, the theoretical reduction factor being very nearly unity.

† That the energy loss in this region is much less than the theoretical value for electrons was proved conclusively by Blackett and Wilson (1937). The reader will find a full account of the experiments and of their relation to a "breakdown of theory" hypothesis in the paper by these authors.

‡ The modification assumed consists in cutting out the electric field of the nucleus inside a certain radius r'—which was the type of hypothesis used before the suggestion of the mesotron. r' is here adjusted to give the observed energy loss by electrons, and then the scattering calculated for this r' using (21). It might be remarked that the rapid variation of r' with energy, which this adjustment requires, constitutes in itself strong *a priori* evidence against the breakdown hypothesis.

is independent of the mass. In fact the results for scattering combined with those for energy loss constitute, independently of other facts, almost irrefutable evidence for the existence of mesotrons.* There is of course considerable other evidence, and the existence of mesotrons is now generally accepted.

The results obtained by Blackett and Wilson for the energy loss indicates that the cosmic-ray particles in the energy range above about 2×10^8 V are nearly all mesotrons. Below this energy the particles observed are mainly electrons as is indicated by their specific ionization and the occurrence of large radiative energy losses characteristic of electronic mass. The results obtained for the scattering therefore generally refer to mesotrons if the energy is above about 2×10^8 V and to electrons if it is below. At very high energies ($\gtrsim 5 \times 10^9$ V) it is also not unlikely that an appreciable proportion are protons.

The expression (37) for the mean deflexion, which is represented by curves A and B, assumes a point nucleus. We have shown (§ 2d) that under the conditions of these cosmic-ray experiments (i.e. $\alpha_{ms} > \bar{\lambda}/b$) the size of the nucleus begins to play an important part, the mean deflexion being actually determined by the ratio of the radius of the atom to that of the nucleus. For a nucleus of radius b the calculated mean deflexion (equ. (40)) is $\bar{\alpha}_b = (19{\cdot}5 - 3{\cdot}1 \log_{10} Z)^{\frac{1}{2}} (\delta)$. This expression and (37) for a point nucleus have practically the same numerical values when $\alpha_{ms} \sim \lambda/b$. Since α_{ms} in the experiments under consideration, though greater than λ/b, is not many times greater, the use of (40) accordingly does not make a big numerical difference.

---

* We say "almost" because it *is* possible to conceive hypotheses that will reduce the energy loss for fast electrons without reducing their scattering. Any such hypothesis however, will be found to be contradicted by other experimental facts, or conflict with well established fundamentals such as the Lorentz transformation. March recently proposed a breakdown of theory hypothesis which in fact would reduce the energy loss without affecting the scattering (though the latter was not intended nor mentioned). In March's theory it is postulated that light quanta of energy ϵmc^2 greater than $h\nu_0 \sim (\hbar c/e^2)\, mc^2 \sim 137mc^2$, cannot be emitted by an electron. Now a fraction $(1 - \epsilon/\xi)$ of the energy lost by a fast electron of initial energy ξmc^2 is lost by the emission of quanta whose individual energy exceeds ϵmc^2. Thus for $\xi > 137$ March's hypothesis reduces the energy loss to about $137/\xi$ of its full theoretical value for electrons, and this would be a fair representation of the observed loss by cosmic-ray particles. The hypothesis leaves the scattering of fast electrons unchanged because the latter is practically independent of the radiative process. It would thus appear a very satisfactory hypothesis. The hypothesis however runs counter to the fundamental tenets of Relativity because it gives the critical frequency ν_0 of the emitted light-quantum the value $(137mc^2/h)$ in the system where the material traversed by the emitting electron is initially at rest, and thereby gives this system a special significance and makes physical laws for it different from those for other systems.

The hypothesis of a critical frequency could be made relativistically invariant if the critical frequency ν_0 is made to refer to the system where the emitting electron is initially at rest. It has however been shown (Williams 1935 *b*) that the emission by a fast electron, in a nuclear collision, of a quantum of energy ϵmc^2 when transformed to a system where the *electron* is initially at rest is merely the emission by the electron of a quantum of energy $(\epsilon/\xi)\, mc^2$. Even for the maximum value of ϵ this is only of the order of mc^2, and is therefore very much less than the critical quantum energy, $h\nu_0$, $\sim 137mc^2$, in March's theory. The application of March's hypothesis in this system (i.e. a "cut-off" at ν_0) would therefore leave the energy loss through the emission of radiation practically unchanged.

The requirements of (40) for electrons and mesotrons respectively are represented by the full curves C and D in fig. 5.* These are in better agreement with the experimental points, which is satisfactory as these curves represent the more correct theoretical result. Actually the average of all the observed points differs by only 3 % from the theoretical value calculated for a finite nucleus, while there is a deviation of about 10 % from the theoretical value for a point nucleus. The statistical error in the experimental value is however of the order of 5–10 %.

There is no evidence in the experiments for scattering by non-electrical forces. Such forces if as strong as the non-electrical forces between slow protons and neutrons (equ. (49)) would increase the average deflexion by about 20 % (equ. (48*a*), footnote, p. 552) and this would bring disagreement with the experimental values.

The scattering arising from a strong short-range interaction with nuclear particles, such as that between slow neutrons and protons, would, as shown in § 2*f*, take place at comparatively large angles where there is practically no scattering due to the electrical forces. Blackett and Wilson plot the distribution of $(E \times \alpha)$ against number. The theoretical distribution of this quantity is independent of E since for fast particles $(\beta \sim 1)$ $\bar{\alpha} \propto \delta \propto 1/E$. Measuring α in degrees, and E in 10^9 V, Blackett and Wilson plot the number of particles scattered by 1 cm. of lead with $E \times \alpha$ in the intervals $0 \rightarrow \pm 0{\cdot}6$, $\pm 0{\cdot}6 \rightarrow \pm 1{\cdot}8, \ldots$. The theoretical distribution of $E\alpha$ for a finite nucleus is practically gaussian with a mean value for 1 cm. of lead as scatterer of $3{\cdot}65\delta \times E, = 0{\cdot}90$ in the above units. The observed and theoretical distributions are plotted in the form of histograms in fig. 6.† The agreement is very satisfactory, the deviations being within the experimental errors estimated by Blackett and Wilson. Thus not only is the average deflexion in good agreement with the theory but also the distribution of the deflexion.

---

* A correction of $(-2\frac{1}{2}\ \%)$ has been applied to equ. (40) for this case, because the condition upon which this equation is based, viz. $\alpha_{ms} \gg \lambda/b$ is not ideally satisfied.

† This figure includes some particles with energy greater than 2×10^9 V $(< 9 \times 10^9)$ which are not included in fig. 5 for the average deflexion. Actually the average deflexion of these particles is somewhat greater than the theoretical value (represented by the continuation of *C–D* in fig. 5). Blackett and Wilson state that the discrepancy may be due to experimental error as the average deflexion, which varies inversely as the energy, is very small for these energies $(\sim \frac{1}{2}^\circ$ at 2×10^9 V). They suggest that if it is real it may be associated with the anomalous energy loss suffered by the same particles.

The feature of the observed distribution which bears on the non-electrical forces is that, of the 116 particles represented, none have $E\alpha > 5$. The theoretical probability of a deflexion giving $E\alpha > 5$ due to electrical interaction is only about 0·0001,* $\sim$0·01 for 116 particles. However, a neutron-proton interaction would make the probability about 0·02,* $\sim$2 for 116 particles, so that the absence of any such deflexions indicates that such interaction does not exist, in conformity with our deductions from the results for the mean deflexion. The cosmic-ray particles concerned are presumably mostly mesotrons and the absence of a proton-neutron-like interaction is in agreement with Yukawa's theory of these particles.† The existing data on the scattering are however inadequate to make any positive

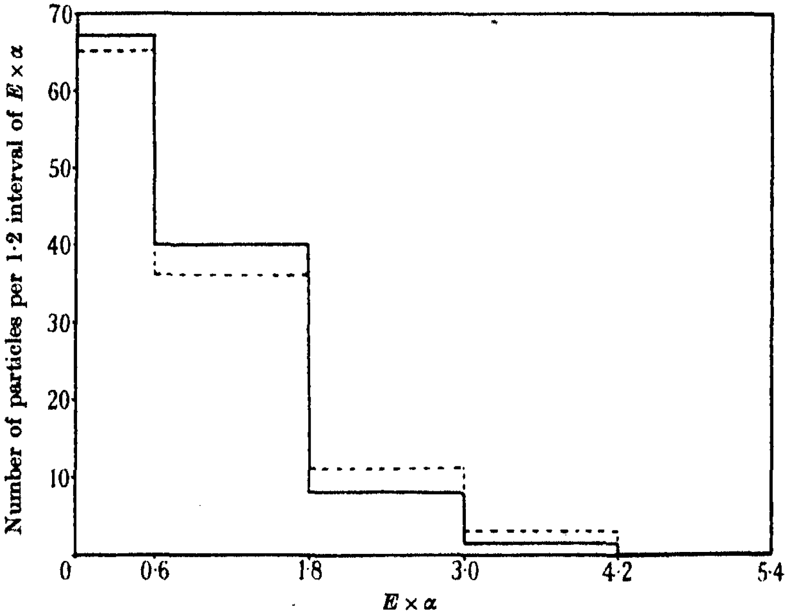


Fig. 6. Distribution of scattering in terms of $E \times \alpha (10^8\ \text{V} < E < 10^{10}\ \text{V})$.

——— Theoretical; - - - - - Observed (Blackett and Wilson).

---

* These estimates are quoted from the end of § 2*f*, the angle $1{\cdot}3\alpha_{ms}$ mentioned there corresponding to $E\alpha = 5$.

† In Yukawa's theory the non-electrical scattering of mesotrons is analogous to the Compton scattering of light-quanta.

pronouncement on the matter, and serve only to show the possibilities of the method.

As bearing on nuclear disintegration by cosmic-ray particles it is of interest to consider the energy transfer corresponding to the maximum observed deflexion. A deflexion through an angle θ means a momentum transfer of $p \times \theta$, where p is the momentum of the scattered particle. If this is taken up by a nuclear proton or neutron, of mass M, assumed to be initially at rest,* the energy transfer, Q, is $p^2\theta^2/2M \sim p^2\phi^2/M \sim 0{\cdot}35E^2\phi^2$ mV. The maximum value of $E\alpha$ in the above observations of Blackett and Wilson is about 4, and this may be taken as the maximum value of $E\phi$, as it is unlikely that the resultant deflexion α after traversing the lead plate is less than the maximum deflexion produced by any one nucleus. $E\phi = 4$ gives $Q \sim 5$ mV. About twice this energy would be required to enable a particle to leave the nucleus. In other words not one of the 116 cosmic-ray particles observed in these experiments is deflected to the extent expected if it had produced a nuclear disintegration in the lead plate, though about ten out of this number would actually pass through a lead nucleus. This would indicate that the "star" disintegrations known to be produced by cosmic-rays, in which several particles are ejected from the nucleus, are unlikely to be caused by fast particles ($\beta \sim 1$) through the kind of interaction which causes the scattering represented in fig. 6. Since protons are known to be present in cosmic-rays these disintegrations may be produced by protons which are sufficiently slow to be stopped by a nucleus and still sufficiently energetic to eject several particles from it.

The low value of the energy transfer to particles of protonic mass, corresponding to the observed scattering, limits the possible interpretations of the anomalous *energy loss* which according to Blackett and Wilson (1937) is suffered by high energy cosmic-ray particles ($\gtrsim 10^9$ V). The average anomalous energy loss is of the order of 100 mV/cm. of lead. If this was spent in nuclear disintegrations in which the energy is taken up by protons or neutrons the scattering would be (from the foregoing calculations) considerably greater than that observed.† If the energy were taken up by electrons the scattering would be negligible but far more electrons would in that case accompany these cosmic-ray particles than actually do so (Wilson 1938). It would indeed appear that to reconcile the anomalous energy loss and the scattering, the former must be assumed to be taken up by particles of intermediate mass such as the recently discovered mesotrons. Creation of slow mesotrons ($\beta \ll 1$) might for example be suggested, or of *neutral* mesotrons if such particles exist.* However, a large cross-section would have to be assumed if the anomalous process takes place in nearly every traversal of a cm. of lead. Further information on the latter and other points is necessary before more definite conclusions can be drawn.

---

* If the nuclear particle moves initially in the direction of the momentum transfer the energy transfer is greater than if it were at rest. This effect is however unlikely to introduce a factor of more than about 2, and of course it can act in the opposite direction, i.e. reduce the energy transfer (cf. Williams 1938 *b*).

† Assuming a loss of 100 mV to take place, and that it is shared amongst n particles of protonic mass, the consequent deflexion of the particle losing this energy would be $\alpha \sim n^{\frac{1}{2}}(100/0{\cdot}35n)^{\frac{1}{2}} E^{-1} \sim 15/E$ degrees, where E is expressed in 10^9 V. Thus for a 10^9 V particle the deflexion would be of the order of 15°, which is vastly in excess of the observed deflexion of these particles.

(*b*) *Scattering of electrons of energy of the order of* 1–20 mV

This has recently been investigated by Fowler and Oppenheimer (1938) and also by Crane and Slawsky (1938). These observers in some of their experiments use foils sufficiently thin that the average angle of scattering is small (~5–10°) so that the calculations of this paper are applicable. The shape of the observed distribution of scattering is in good agreement with that to be expected, viz. a gaussian curve with a single-scattering "tail". The average angle of scattering in the experiments of Crane and Slawsky on the scattering of 0·9 mV electrons by an aluminium foil 0·0025 cm. thick, corresponds to an average deflexion of 5·5° ± 0·5. The theoretical value according to (37)† is 6·4°, or 5·9° after allowing for the exclusion of certain angles of scattering by the experimental arrangement (see Appendix D). The agreement is thus fairly satisfactory. The results obtained by Fowler and Oppenheimer for the scattering of higher energy electrons (~10 mV) by lead foils are, however, in disagreement with the theoretical values. The observed average deflexion is only about one-half the theoretical.

---

* There is experimental evidence, and also theoretical reasons, for the spontaneous transformation of a mesotron to an electron and a neutrino, with a period for a slow mesotron of about 10^{-6} sec. Accordingly the creation of slow mesotrons which are stopped in the lead plate (mesotrons are hardly ever observed to emerge from the plate so that the creation of fast mesotrons is ruled out) would give rise almost simultaneously to fast electrons with energy of the order of 100 mV. The absence of any appreciable number of such secondary electrons would thus exclude the creation of charged mesotrons.

† The conditions here are rather far removed from those for the scattering by a 1 cm. lead plate and the expression for the average angle of scattering given (37*a*) (footnote *, p. 550) is probably more accurate than (37). The difference between them is however small (37*a*) giving $\bar{\alpha} = 6{\cdot}3°$ (after correcting for the experimental arrangement). It might be noted that the theoretical contribution from the single scattering tail is only 0·2°. This tail is actually not represented in the observations but this is not necessarily in conflict with the theory since with the limited number of electrons observed and the "cut-off" imposed by the experimental arrangement slightly less than 1 electron would on the average be expected to be deflected into this tail.

The details of the relation of theory and experiment will appear in a paper by Fowler (1938). A factor of $\frac{1}{2}$ in the average angle of scattering corresponds to a factor of $\frac{1}{4}$ in the intensity of single scattering, the angles involved ranging from $\phi_{\min} \sim \lambda/a$ to $\phi_2 \sim 200\phi_{\min}$. Alternatively it may be interpreted as a complete "cut-out" of the scattering for angles between ϕ' and α_{ms} with no reduction in the scattering for angles $< \phi'$. Then to fit the observations $\log(\phi'/\phi_{\min}) \sim \frac{1}{4}\log(\phi_2/\phi_{\min})$.* This gives $\phi' \sim 10\phi_{\min}$, corresponding to a distance of approach to the nucleus of $a/20 \sim 2(\hbar/mc)$. The experimental results would thus indicate that the effective nuclear field inside this radius is much less than the Coulombian field. Indeed, indications of such abnormalities have also come from the study of the single scattering of much lower energy electrons (~ 1 mV)†: for instance, according to recent experiments by Barber and Champion (1938), single scattering of one-million-volt electrons by mercury through angles $\gtrsim 20°$ is only about 15 % of the theoretical. The theoretical distance of approach in such collisions ($\sim \lambda/\theta$) extends out to about $1{\cdot}5(\hbar/mc)$, and is therefore of the same order of magnitude as it is in the collisions cut out to account for the results of Fowler and Oppenheimer. It is very difficult to believe in these interpretations, particularly as the theoretical scattering is in good agreement with experiment for electrons of about 10^8 V energy (as shown by the experiments of Blackett and Wilson for cosmic-rays), and as there is no indication at all of anomalies in the behaviour of the K-electrons of heavy elements, which spend a large part of their time within distances of the order of $\hbar/mc$ of the nucleus. If the experimental results are confirmed they will require some radical innovations in the theory.

The calculations described in this paper were carried out at Liverpool University, and I have great pleasure in thanking Professor J. Chadwick for his encouraging interest in this and allied work. I also wish to thank Professor P. M. S. Blackett and Dr J. G. Wilson for the opportunity of several discussions with them on the problem.

APPENDIX A

Effect of shielding. We first have to evaluate $\int_0^{\phi_1} \theta^2 P(\theta)\, d\theta$ (see § 2*b*). $P(\theta) = (2k/\theta^3) \times F(2a \sin \frac{1}{2}\theta/\lambda) = (2k/\theta^3) \times F(a\theta/\lambda)$. $(2k/\theta^3)$ represents the intensity of scattering through small angles for an unshielded nucleus.

---

* Cf. footnote *, p. 550.

† Other serious discrepancies between theory and experiment have been found in this region, such as a much greater scattering by nitrogen than is required by theory (Stepanowa 1937) and too frequent pair production (Skobelzyn 1934).

F represents the effect of shielding (§ 1 *b*, equ. (19)), and since from (32*a*) $\phi_1 \gg \phi_{\min} \approx (\bar{\lambda}/a)$ F is unity in the region of ϕ_1. The above integral may be written

$$\int_0^{\phi_1} \theta^2 (2k/\theta^3)\, F(a\theta/\bar{\lambda}) = 2k \int_{\theta=0}^{\phi_1} F(a\theta/\bar{\lambda})\, d \log (a\theta/\bar{\lambda}).$$

In fig. 7 the values of $F(a\theta/\bar{\lambda})$ calculated by Bullard and Massey (1930) are plotted against log $(a\theta/\bar{\lambda})$. The approximation previously made (§ 1 *a*, and equ. (30*b*)) corresponds to taking $F = 1$ for $\theta > \bar{\lambda}/a$, and $F = 0$ for $\theta < \bar{\lambda}/a$. It will be seen that this somewhat underestimates the rate at which shielding reduces the scattering for small angles. A numerical integration of the area under the curve in fig. 7 shows that the effective lower limit $\theta_{\min}$—defined by $\int_0^{\phi_1} F(a\theta/\bar{\lambda})\, d \log \theta = \log (\phi_1/\theta_{\min})$—is $2{\cdot}10\bar{\lambda}/a$. Thus

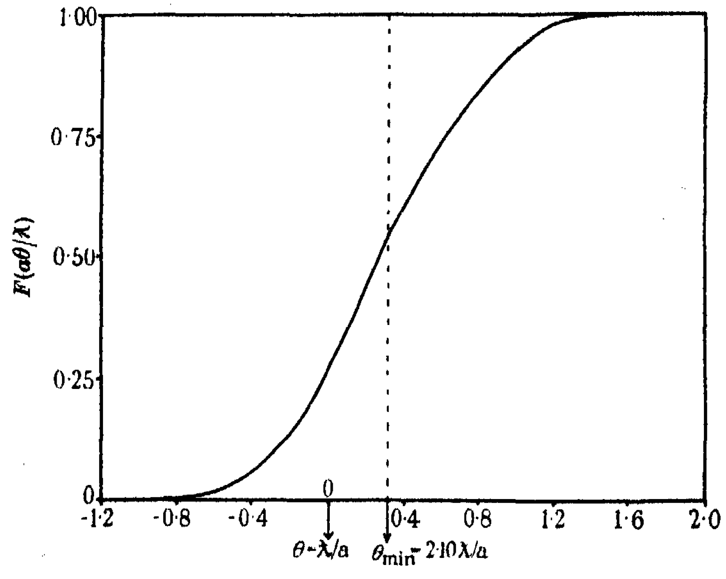


FIG. 7. Effect of shielding.

$$\int_0^{\phi_1} \theta^2 P(\theta)\, d\theta = 2k \log (\phi_1/2{\cdot}10\overline{\bar{\lambda}/a}). \tag{i}$$

In fig. 8 the angle θ is represented by a radius vector in the $\phi\phi'$ plane, ϕ and ϕ' being the projected angles of scattering on two mutually perpendicular planes intersecting along a line parallel to the initial direction. Since the angles are small $\theta^2 = \phi^2 + \phi'^2$. The scattering is symmetrical about the initial direction and therefore on the average $\phi^2 = \phi'^2 = \frac{1}{2}\theta^2$. It follows that the contribution of the scattering through $\theta < \phi_1$ to $\int_{\phi=0}^{\phi_1} \phi^2 P(\phi)\, d\phi$ is equal to $\frac{1}{2}\int_{\theta=0}^{\phi_1} \theta^2 P(\theta)\, d\theta = k \log(\phi_1/2{\cdot}10\bar{\lambda}/a)$. This however is not the full value of $\int_{\phi=0}^{\phi_1} \phi^2 P(\phi)\, d\phi$. We must add the effect of scattering through angles $\theta > \phi_1$ for which $\phi < \phi_1$. Such θ are represented by the shaded area in fig. 8. Since in this area $\theta > \phi_1 \gg \bar{\lambda}/a$ the distribution of θ is given by the "unshielded" formula, i.e. $P(\theta) = 2k/\theta^3$. Accordingly the probability that ϕ is between ϕ and $\phi + d\phi$ and ϕ' between ϕ' and $\phi' + d\phi'$ is

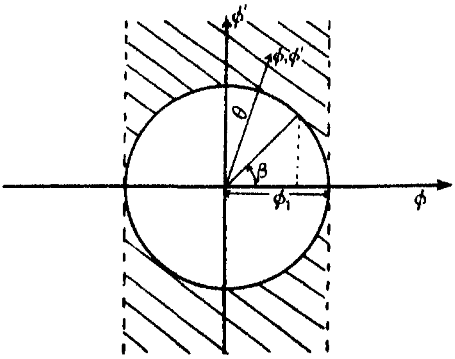


FIG. 8

$$\{(2k/\theta^3) \div 2\pi\theta\} \times d\phi\, d\phi' = (k/\pi)(\phi^2 + \phi'^2)^{-2}\, d\phi\, d\phi',$$

and the contribution of the shaded area to $P(\phi)\, d\phi$ is

$$P'(\phi)\, d\phi = 4\int_{\phi' = (\phi_1^2 - \phi^2)} (k/\pi)(\phi^2 + \phi'^2)^{-2}\, d\phi\, d\phi'.$$

The integral converges so rapidly, and the θ concerned are so small, that the upper limit may be taken as infinite. The factor 4 allows for the four quadrants in fig. 8, $P(\phi)$ referring to the absolute value of ϕ. The integration gives

$$P'(\phi) = (2k/\pi)\,\phi^{-3}(\tfrac{1}{2}\pi - \beta - \tfrac{1}{2}\sin 2\beta), \tag{ii}$$

where $\cos\beta = \phi/\phi_1$ (see figure). For $\beta = 0$ $P'(\phi)$ reduces to k/ϕ^3, which proves formula (27) for the distribution of the projected deflexion ϕ due to all θ (when there is no shielding). The contribution of the shaded area to

$\int_0^{\phi_1}\phi^2 P(\phi)\,d\phi$ is now obtained by integrating $\phi^2 P'(\phi)$ from $\phi = 0$ to $\phi = \phi_1$, i.e. from $\beta = \frac{1}{2}\pi$ to $\beta = 0$. The contribution is rather small, viz. $0\cdot184k$. Adding it to the main contribution, viz. $k\log(\phi_1/2\cdot10\overline{\bar{\lambda}/a})$, gives $k\log(\phi_1/1\cdot75\overline{\bar{\lambda}/a})$. The effective lower limit $\phi_{\min}$ to ϕ due to shielding, defined by $\int_0^{\phi_1}\phi^2 P(\phi)\,d\phi = \int_{\phi_{\min}}^{\phi_1}(k/\phi^3)\,\phi^2\,d\phi$, is thus

$$\phi_{\min} = 1\cdot75(\bar{\lambda}/a) \tag{iii}$$

and

$$\bar{\alpha}_1^2 = (2/\pi)\,k\log(\phi_1/1\cdot75\overline{\bar{\lambda}/a}). \tag{iv}$$

In terms of δ (equ. (31)) as unit of angle, and substituting for ϕ_1, $\bar{\lambda}$, and a, (iv) may be written

$$\overline{\alpha_1}^2 = \log(Z^{\frac{4}{3}}Nt\hbar^2/1\cdot75^2 2\pi m^2\beta^2c^2) = \log M, \text{ say.} \tag{iv$a$}$$

Appendix B

Effect of scattering through $\phi > \phi_1$. ϕ_1 is defined such that on the average one collision takes place with a projected deflexion $\phi > \phi_1$, when the particle passes through the plate. We require the extent to which this deflexion increases the arithmetic mean projected deflexion, the value of which ($\bar{\alpha}_1$) due to $\phi < \phi_1$ is given by (36) or equ. (iv) of Appendix A. We shall denote the increase by $\bar{\alpha}_2$.

From a well-known theorem in probability a thing which happens once on the average does not happen at all in a fraction e^{-1} ($\sim 0\cdot37$) "trials", happens once in a fraction e^{-1}, ..., happens n times in a fraction $(e^{-1}/n!)$, ... Thus in a fraction e^{-1} traversals of the scattering plate a collision giving $\phi > \phi_1$ does not happen at all. On such occasions the resultant deflexion, which we shall denote by α, is just the deflexion α_1 arising from the scattering through $\phi < \phi_1$. The average value of α for these traversals is therefore the same as the average value of α_1 given by (36), and the distribution of α is the same as the distribution of α_1 given by (29).

A deflexion through $\phi > \phi_1$ takes place *once* in a fraction e^{-1} traversals. We shall denote by α_2 the resultant deflexion due to $\phi > \phi_1$. When there is only one such deflexion the distribution $P_2(\alpha_2)$ of α_2 is evidently the same as the single scattering distribution, i.e. $P_2(\alpha_2) = k/\alpha_2^3$, provided $\alpha_2 > \phi_1$. For $\alpha_2 < \phi_1$, $P_2(\alpha_2)$ is of course zero. The problem now is to calculate the increase in the average deflexion when this distribution $P_2(\alpha_2)$ is superimposed on the gaussian distribution $P_1(\alpha_1)$ of α_1. A general formula for this increase may readily be obtained. Let us consider those traversals for which

$\alpha_1 = \alpha_1'$, $\alpha_2 = \alpha_2'$. Then if $\alpha_1' > \alpha_2'$ the resultant deflexion is always in the direction of α_1', and its arithmetic mean value is α_1', since α_2' is as likely to be in the same as in the opposite direction to α_1'. Similarly, if $\alpha_1' < \alpha_2'$ the resultant deflexion is always in the direction of α_2', and the average is α_2'. A deflexion α_2' therefore increases the average deflexion only when $\alpha_2' > \alpha_1'$ and the average increase is then equal to $\alpha_2' - \alpha_1'$. It follows that the increase in the average deflexion due to $\alpha_2 = \alpha_2'$, for all α_1, is

$$\gamma_2 = \int_{\alpha_1=0}^{\alpha_1=\alpha_2'} P_1(\alpha_1)\,(\alpha_2'-\alpha_1)\,d\alpha_1.$$

The average increase due to all α_2 is then

$$\bar{\gamma}_2 = \int_{\alpha_2=0}^{\infty} P_2(\alpha_2)\,\gamma_2(\alpha_2)\,d\alpha_2 = \int_{\alpha_2=0}^{\infty} P_2(\alpha_2)\,d\alpha_2 \int_{\alpha_1=0}^{\alpha_2} P_1(\alpha_1)\,(\alpha_2-\alpha_1)\,d\alpha_1. \quad \text{(v)}$$

The resultant average deflexion $\bar{\alpha}$ is

$$\bar{\alpha} = \bar{\alpha}_1 + \bar{\gamma}_2 = \int_{\alpha_1=0}^{\infty} P_1(\alpha_1)\,\alpha_1\,d\alpha_1 + \int_{\alpha_2=0}^{\infty} P_2(\alpha_2)\,d\alpha_2 \int_{\alpha_1=0}^{\alpha_2} P_1(\alpha_1)\,(\alpha_2-\alpha_1)\,d\alpha_1, \quad \text{(v}a\text{)}$$

$$\bar{\alpha} = \bar{\alpha}_2' + \bar{\gamma}_1 = \int_{\alpha_2=0}^{\infty} P_2(\alpha_2)\,\alpha_2\,d\alpha_2 + \int_{\alpha_1=0}^{\infty} P_1(\alpha_1)\,d\alpha_1 \int_{\alpha_2=0}^{\alpha_1} P_2(\alpha_2)\,(\alpha_1-\alpha_2). \quad \text{(v}b\text{)}$$

(vb) is equivalent to (va), differing from it only in the order of integration.

$\bar{\gamma}_2/\bar{\alpha}_1$ depends on $\bar{\alpha}_2'/\bar{\alpha}_1$, and this depends on the thickness and nature of the scattering material. However as this dependency is through a log term, and therefore not a sensitive one, $\bar{\gamma}_2/\bar{\alpha}_1$ changes very little even though there is a large variation of these factors. Actually if $\bar{\gamma}_2$ is calculated for a typical case the value for all practical variations from that case may easily be deduced.

We shall take the scattering of fast particles ($\beta = 1$) by a plate of lead 1 cm. thick as standard. The value of $\bar{\alpha}_1$ (equ. (36)) in terms of δ as unit is then 3·23. Substituting this in (29) and carrying out the double integration in (iv) we find (in units δ)*

$$\bar{\gamma}_2 = 0{\cdot}88, \quad \text{(vi)}$$

$$d\bar{\gamma}_2/d\bar{\alpha}_1 = -0{\cdot}21. \quad \text{(vii)}$$

---

* The integration does not involve much numerical work. For $\alpha_2 \gtrsim 12$, $\sim 4\bar{\alpha}_1$, the second integral in (v) is to a few parts in a 1000 equal to $\alpha_2 - \bar{\alpha}_1$, and the corresponding contribution to $\bar{\gamma}_2$ can then be calculated analytically. For $\alpha_2 < 12$ the exponential in (29) was expanded in powers of α_1^2, a simple integration then giving the corresponding contribution to $\bar{\gamma}_2$ in a power series. For the α_2 concerned this series converges fairly rapidly, six terms giving the result to less than 1 in a 1000. To check the numerical work $\bar{\gamma}_2$ was also calculated using (vb), thus reversing the order of integration. The same result was obtained.

Of the above increase due to $\phi > \phi_1$ about 0·20 comes from the single scattering tail beyond $\alpha \sim \phi_2$ (see fig. 1); the remainder, $\sim 0{\cdot}70$, represents a broadening of the gaussian distribution.

Let us next consider the traversals in which two collisions giving $\phi > \phi_1$ take place. These represent $(e^{-1}/2!) \sim 0{\cdot}18$ of the total number of traversals. The first collision superimposed on the gaussian distribution (29) will increase the average deflexion by 0·88. The second will not produce an additional increase of quite the same amount because the first collision has made the distribution a little broader than the initial gaussian distribution (29). 0·88 is therefore an upper limit to the effect of the second collision. A lower limit, but a poor one, is zero. Similarly the upper limit to the effect of three collisions is $3 \times 0{\cdot}88$ and a lower limit is 0·88, and for n collisions the limits are $n \times 0{\cdot}88$ and 0·88. Remembering that n collisions through $\phi > \phi_1$ occur only in a fraction $(e^{-1}/n!)$ traversals we find as an upper limit to the average effect of $\phi > \phi_1$ for *all* traversals

$$\bar{\alpha}_{2\max} = e^{-1}\left(1 + \frac{2}{2!} + \frac{3}{3!} + \ldots + \frac{n}{n!} + \ldots\right) \times 0{\cdot}88 = 0{\cdot}88,$$

and as lower limit

$$\bar{\alpha}_{2\min} = e^{-1}\left(1 + \frac{1}{2!} + \frac{1}{3!} + \ldots + \frac{1}{n!} + \ldots\right) \times 0{\cdot}88 = 0{\cdot}56.$$

Adding this to $\bar{\alpha}_1 = 3{\cdot}23$ we get a resultant, $\bar{\alpha}$, between 3·79 and 4·11, the mean of these two values being therefore certainly correct within about 3 %.

The true value of $\bar{\alpha}_2$ is evidently much nearer the upper limit, and an estimate of a higher order of accuracy may be made in the following way. The average value, $\bar{\phi}$, of ϕ for single scattering through $\phi > \phi_1$ is $\int_{\phi_1}^{\infty} (\pi/\phi^3)\,\phi\,d\phi = \sqrt{(2\pi)}$. If the distribution of ϕ giving this average was gaussian its superposition on the gaussian distribution of α_1, represented by (29), would increase the average deflexion by

$$\bar{\gamma}_2 = (\bar{\alpha}_1^2 + \bar{\phi}^2) - \bar{\alpha}_1 = (3{\cdot}23^2 + 2\pi) - 3{\cdot}23 = 0{\cdot}85.$$

The close agreement with (vi) shows that for a given value of $\bar{\alpha}_1$ and $\bar{\phi}$ the resultant average deflexion is little dependent on the detailed forms of the distribution of α_1 and ϕ. This procedure may therefore be safely used to estimate the slight difference between the effect of two collisions through $\phi > \phi_2$ and twice the effect of one, and between the effect of three and three times the effect of one.* The method gives $\bar{\alpha}_2 = 0{\cdot}82$, which is 0·06 less than the upper limit of 0·88. This then is the final estimate of the increase in the average deflexion due to single scattering through $\phi > \phi_1$. The average deflexion due to $\phi < \phi_1$ is $\bar{\alpha}_1 = 3{\cdot}23$, so that collisions with $\phi > \phi_1$ increase the average deflexion by about 25 %. The arithmetic mean deflexion due to all angles of scattering is $\bar{\alpha}_1 + \bar{\alpha}_2 = 4{\cdot}05$. This applies to the scattering by 1 cm. of lead. The distribution of single scattering for $\phi > \phi_1$ in the units used in these calculations is independent of the conditions, being represented by π/ϕ^3. The effect of a departure from the conditions of 1 cm. of lead and $\beta = 1$ is to change $\bar{\alpha}_1$, i.e. the width of the gaussian distribution $P_1(\alpha)$. $\bar{\alpha}_1$ is given by (36), and in the units δ it is not sensitive to the conditions owing to its logarithmic form. A change in the thickness of the scattering plate from 1 cm. to 0·01 cm., for instance, reduces $\bar{\alpha}_1$ by only about 20 %. The smaller $\bar{\alpha}_1$ is, the greater is the relative effect of the scattering through $\phi > \phi_1$, i.e. the greater is $\bar{\alpha}_2$. For 1 cm. of lead and $\beta = 1$ we find on the basis of (vii) $\delta\bar{\alpha}_2 = -0{\cdot}20\delta\bar{\alpha}_1$, and therefore

$$\begin{aligned}\delta\bar{\alpha} = \delta(\bar{\alpha}_1 + \bar{\alpha}_2) = 0{\cdot}80\delta\bar{\alpha}_1 &= 0{\cdot}80\,\delta(\log M)^{\frac{1}{2}} \text{ (from (iv}a\text{), Appendix A))}\\ &= 0{\cdot}40\,\delta(\log M) \div \bar{\alpha}_1 = 0{\cdot}123\log M/M_0,\end{aligned}$$

where M_0 denotes the value of M for 1 cm. of lead. This expression for $\delta\bar{\alpha}$ (which is equivalent to the first term in a Taylor expansion) is accurate within about 1 % for a 100-fold variation of M, and is therefore adequate for our purpose. The arithmetic mean deflexion in the general case is therefore $= 4{\cdot}05 + 0{\cdot}123\log_e(M/M_0) = 4{\cdot}05 + 0{\cdot}28\log_{10} M/M_0$. The unit is δ and introducing this and substituting for M/M_0, we have

$$\bar{\alpha} = \{3{\cdot}69 + 0{\cdot}28\log_{10}(Z^{\frac{4}{3}}\rho t/W\beta^2)\}\,(\delta), \tag{viii}$$

which is equ. (37) of the text.

The above method has also been used to calculate the average deflexion, $\bar{\alpha}'$, when the effect of single scattering through angles greater than ϕ' is left out, for various values of ϕ'. It is found that if ϕ' is less than ϕ_2 (equ. (32), fig. 1) $\bar{\alpha}'^2$ is given within a few per cent by $\int_{\phi_{\min}}^{\phi'} \phi^2 P(\phi)\,d\phi$. This result is used in the text to derive equ. (40) for the average deflexion allowing for the finite size of the nucleus. The latter greatly reduces the single scattering formula for $\phi \gtrsim \lambda/b$, so that ϕ' corresponds to λ/b.

---

* Further accuracy may be achieved if this procedure is adopted only for the broadening of the gaussian distribution ($\phi < \phi_2$), and the effect of n collisions through $\phi > \phi_2$ then taken as n times the effect of one. The difference this makes in the results is however insignificant.

APPENDIX C

Computational errors in derivation of mean deflexion, etc. Provided the angle of scattering is small the computational error in the result (37) is less than 1 %. The assumption of small angles introduces an error into the formula (27) for the probability of a projected deflexion ϕ of the order of only ϕ^4 (+higher terms), and is therefore negligible. An error of a higher order arises from the fact that the "initial" direction for single scattering changes as the particle traverses the plate. If we denote by α' the component of the resultant deflexion on the plane perpendicular to that containing the observed projected deflexion α, then for such a traversal the correcting factor for this error is $\sec\alpha' \sim (1+\frac{1}{2}\alpha'^2)$, which on the average $= (1+\frac{1}{4}\pi\bar{\alpha}^2)$. The scattering of the particle in the plate also increases the effective thickness of material traversed. This increases the average angle of scattering by $(1+\frac{1}{8}\pi\bar{\alpha}^2)$. Combining the two gives a correcting factor $(1+\frac{3}{8}\pi\bar{\alpha}^2)$. This estimate of course also assumes that α is small, but it is sufficient to give a rough idea of the correction for $\bar{\alpha}$ up to about 30°. In the experiments considered in this paper $\bar{\alpha}$ varies from a fraction of a degree to about 10°, and even for the latter the correction is only about 4 % and may be discarded. We have also estimated the effect of the terms $(\pi\alpha\beta^2\sin\frac{1}{2}\theta)$ and $(-\beta^2\sin^2\frac{1}{2}\theta)$ in Mott's relativistic formula (equ. (14*b*)), which are neglected in the derivation of (37). While these terms affect quite appreciably the single scattering through large angles $\sim 90°$, their effect in the region of multiple scattering in the experiments under consideration here can be neglected.

Lastly we must consider the possible error arising from our assumption that the distribution of multiple scattering due to single scattering through $\phi < \phi_1$ is exactly gaussian (equ. (29)), with an average deflexion $\bar{\alpha}_1$ given by (30). The essential condition for this is that the average number of collisions in any range of ϕ which contributes appreciably to the average resultant deflexion, $\bar{\alpha}$, be large. Owing to the rapid decrease in the number of collisions with increasing ϕ this is least satisfied for the largest angles involved, i.e. ϕ near ϕ_1. Actually the average number of collisions which give ϕ between $\frac{1}{2}\phi_1$ and ϕ_1 is 3, while they contribute $(\frac{1}{2}\log 2)/\log(\phi_1/\phi_{\min} \sim 300) \sim 6$ % to $\bar{\alpha}_1$. Now this contribution may be obtained with higher accuracy by assuming a gaussian distribution for the multiple scattering due to $\phi < \frac{1}{2}\phi_1$ and using the general formula (35) to get the contribution from $\frac{1}{2}\phi_1 < \phi < \phi_1$. The result shows that the expression (30) for $\bar{\alpha}$ represents this contribution with an error of about 0·3 %, corresponding to an error of only 0·3 % × 6 % $\sim$ 0·02 % in $\bar{\alpha}$. The error arising from all other regions of ϕ (i.e. $\phi < \frac{1}{2}\phi_1$) is of the same order, so that the error in $\bar{\alpha}$ as given by (30) is only a few parts in 10,000 and can therefore be neglected. The departure of the *shape* of the distribution due to $\phi < \phi_1$ from a gaussian form is similarly negligible.

An important point to be mentioned here is that though the regions $\phi < \phi_1$ and $\phi > \phi_1$ have been treated separately in our calculations the contribution of different ϕ to $\bar{\alpha}$ fulfils satisfactorily the condition of continuity at $\phi = \phi_1$. The slight discontinuity that exists is only about $(1/6)\,\phi_1^2/\pi\bar{\alpha}_1^2$, $\sim 0{\cdot}005$ for the standard case considered (1 cm. lead, $\beta = 1$). The choice of angle for dividing the two regions is of course governed only by the convenience of the calculations. A division at an angle ϕ' differing appreciably from the angle ϕ_1 chosen here would make the accurate treatment of $\phi > \phi'$ very laborious if $\phi' \ll \phi_1$. On the other hand if we chose $\phi \gg \phi'$ the gaussian representation of the effect of $\phi < \phi'$ would not be valid. The situation is much the same as in the problem of the distribution of ionization energy losses by fast particles in passing through thin foils (Williams 1929) where collisions giving an energy loss $Q > S$ and those giving $Q < S$ are treated separately in the above fashion, S being such that on the average one collision takes place with $Q > S$.

Appendix D

Limitation of range of measured angles of scattering imposed by experimental arrangement. In the cloud chamber experiments on scattering such as those referred to in this paper angles of scattering greater than a certain value are automatically excluded from the observations. In the experiments of Crane and Slawsky, for example, only those tracks are included which have a length greater than 7·5 cm. on the emergent side of the scattering foil. The depth of the illuminated part of the chamber in the direction of photography is about 2·5 cm. while the scatterer is placed in the middle of the illuminated region. Thus, as Slawsky (1938) points out the component, α', of the resultant deflexion in the direction of photography is limited to $(2{\cdot}5/2 \times 7{\cdot}5) \sim 10°$. This component is in the same direction as the component ϕ' of fig. 8, and in the plane perpendicular to that containing the measured projected deflexion α. In the cosmic-ray experiments of Blackett and Wilson α' is limited by the counter control, its maximum value, which we shall denote by ψ, being equal to half the sum of the angles subtended by the width of the scattering plate at the two controlling counters (which are placed above and below the chamber respectively). The finite size of the counters increases ψ somewhat. When these limitations are of importance it would appear to be a better procedure to calculate the theoretical scattering for the conditions of the experiment, rather than attempt to correct for the excluded scattering from a knowledge of the scattering in the observed region (as is done by Slawsky). We shall therefore consider briefly the theoretical scattering when α' is limited to values less than ψ. Since throughout the paper we have only concerned ourselves with small angles of scattering we need only consider the effect when ψ is small ($< 20°$ say). (When ψ is large the probability of scattering through small angles is evidently unaffected.)

In the region of multiple scattering where the distribution of α and α' are gaussian the probability of α having a given value is independent of the value of α'. The observed shape of the distribution of α is therefore unaffected by a limitation of α' to $\alpha' < \psi$.* It is only where the scattering is "single" (and to a less extent where it is plural) that the limitation of α' affects the distribution of α. We shall consider this effect when the single scattering is given by the Rutherford-like formula (13), which for small angles gives $P(\alpha) = k/\alpha^3$ (equ. (27)). The latter assumes that there is no limitation of the component α'. If this is limited to $\alpha' < \psi$ then from equ. (ii), Appendix A,

$$P(\alpha), \quad = P_\psi(\alpha), \quad = (2/\pi)(k/\alpha^3)(\beta + \tfrac{1}{2}\sin 2\beta), \qquad \text{(ix)}$$

where $\tan\beta = \psi/\alpha$.

To a first approximation the effect of the limitation of α' may now be allowed for if we use (ix) in considering the single scattering tail instead of (27). Numerical evaluation of (ix) shows that $P(\alpha)$ is reduced by about 20 % at $\alpha = \psi$, and to 50 % at $\alpha = 2\psi$. In the experiments of Crane and Slawsky $\psi = 10°$, so that single scattering which starts at about 20° is reduced to about a half at that point. The single scattering tail is thus appreciably reduced. As this tail contributes very little to the average deflexion, $\bar{\alpha}$, the effect on the latter is small. Actually it is found to be reduced from 6·4° to 5·9°.†

In the cosmic-ray experiments of Blackett and Wilson the effect is negligible except possibly for the single scattering tail for the low energy particles observed ($\sim 10^8$ V). The limitation of the component α by the experimental arrangement usually takes place at much larger angles than does the limitation of the component α_1' and the effect can be neglected in all the experiments considered here. These effects however should be borne in mind when experiments of this kind are extended.

---

* It is to be noticed that the experimental arrangement does not of course impose any limitation on the deflexions produced by different nuclei *inside* the foil. Deflexions with $\phi' > \psi$ may take place and the particle still be included if the component α' of the resultant deflexion on emergence is less than ψ.

† Let P_{ex} = probability (theoretical) that a track is excluded from the observations (due to $\alpha' > \psi$), and let $\bar{\alpha}_{\text{ex}}$ = average α for excluded tracks, α_{in} = average for included tracks, $\bar{\alpha}$ = average for all tracks. Then $\bar{\alpha} = \alpha_{\text{in}}(1 - P_{\text{ex}}) + \alpha_{\text{ex}} P_{\text{ex}}$. When the difference between α and α_{in} is large this equation, or its equivalent, should be used to calculate the difference, in order to avoid any errors due to incorrect normalization.

SUMMARY

The relativistic theory of nuclear scattering through small angles is discussed, and corrections are applied for the effects of the finite size of the nucleus and the shielding of the nucleus by the atomic electrons. This is preliminary to calculations on the multiple scattering of fast electrons by thin metal foils and of cosmic-ray particles by metal plates. These calculations lead to an accurate estimate of the average theoretical deflexion under the conditions of the experiments, and thereby permit a quantitative interpretation of the observations to be made. A general account of the statistical calculations, and an approximate treatment, are given in § 2*a*.

The experimental results of Blackett and Wilson on the scattering of cosmic-ray particles are in good agreement with the theoretical values, especially if the effect of the finite size of the nucleus is allowed for. The evidence which this agreement provides for the existence of mesotrons is discussed. There is no positive evidence in the experiments for a short-range non-electrical force between the cosmic-ray particles and the nuclear particles of the matter traversed, similar to that between slow neutrons and protons. It is also inferred that the particles whose scattering was observed produced no nuclear disintegrations in the scattering plates. It is suggested that the observed anomalous loss of energy by cosmic-ray particles with energy around 10^9 V may be due to the production of neutral mesotrons, if such particles exist.

The scattering of fast electrons (~ 10 mV) by thin lead foils as recently observed by Fowler and Oppenheimer is much less than the present theoretical values. This discrepancy seems to fall in line with observations by Klarman and Bothe and by Champion on the single scattering in heavy gases of electrons with energy of the order of 1 mV. The scattering of million-volt electrons by a thin aluminium foil recently observed by Crane and Slawsky is in fair agreement with the present calculations, so that the above discrepancy if real must be associated with a high atomic number. There is as yet no indication of how the discrepancy could be accounted for theoretically.

REFERENCES

Anderson 1933 *Phys. Rev.* **43**, 381.
Barber and Champion 1938 *Proc. Roy. Soc.* A, **168**, 159.
Blackett and Wilson 1937 *Proc. Roy. Soc.* A, **160**, 304.
— — 1938 *Proc. Roy. Soc.* A, **165**, 209.
Bothe 1929 *Z. Phys.* **54**, 161 (also earlier papers).
Bullard and Massey 1930 *Proc. Camb. Phil. Soc.* **26**, 556.
Crane and Slawsky 1938 *Phys. Rev.* (in preparation).
Fowler 1938 *Phys. Rev.* (in the Press).
Fowler and Oppenheimer 1938 *Phys. Rev.* **54**, 320
Heisenberg 1937 *Leipzig. Akad. der Wissen.* **89**, 369.
Klarman and Bothe 1936 *Z. Phys.* **101**, 489.
March 1937 *Z. Phys.* **108**, 128.
Mott 1929 *Proc. Roy. Soc.* A, **124**, 425.
Skobelzyn 1934 *Nature, Lond.*, **133**, 646.
Slawsky 1938 Thesis submitted for Ph.D. University of Michigan.
Stepanowa 1937 *Phys. Z. Sowjet.* **12**, 5, 550.
Wentzel 1922 *Ann. Phys., Lpz.*, **69**, 335.
Williams 1929 *Proc. Roy. Soc.* A, **125**, 420.
— 1935*a* *Phys. Rev.* **47**, 568. Also *Kern Physik*, p. 123. Berlin: Springer.
— 1935*b* *K. danske vidensk. Selsk.* **13**, 36.
— 1938*a* *Phys. Rev.* **53**, 433.
— 1938*b* *Nature, Lond.*, **142**, 131.
Wilson 1938 *Nature, Lond.*, **142**, 72.

The diffusion of radon gas mixtures

By W. Hirst, B.Sc. and G. E. Harrison, Ph.D.
Physics Department, University of Birmingham

(*Communicated by M. L. E. Oliphant, F.R.S.—Received* 28 *October* 1938)

The repulsive force, F_{12}, between dissimilar molecules may be represented by the equation

$$F_{12} = \mu_{12} d^{-s_{12}},$$

where d is the distance between the dissimilar molecules which are regarded as point centres of repulsive force, while s_{12} represents the repulsive force index, and μ_{12} the repulsive force constant between these molecules.

Chapman (1929) has given a special theory of thermal diffusion in mixtures of two gases in which one component is rare and heavy. He has shown that F_{12} may be determined provided that thermal diffusion measurements for the given gas mixture have been made, and the coefficient of viscosity of the lighter gas at various temperatures together with the coefficient of diffusion of the two gases is known.

Thermal diffusion measurements for radon-hydrogen and radon-helium gas mixtures have already been made by one of us (Harrison 1937), radon being the rare and heavy constituent of these mixtures. The variation of the viscosity of hydrogen and helium over a wide range of temperature was known, but no experimental value of the coefficient of diffusion of radon into hydrogen or helium was available. In the previous investigation, approximate values of the latter were therefore calculated from a formula given by Chapman (1917). Owing to uncertainty in the accuracy of these calculated values, s_{12} was evaluated approximately for radon-hydrogen and radon-helium molecules, but μ_{12} was not calculated, since the latter is much more sensitive to an error in the coefficient of diffusion than the former. The purpose of the present investigation was to obtain experimental values of the coefficient of diffusion of radon into hydrogen, helium, neon and argon, so that the repulsive force between radon-hydrogen and radon-helium molecules can be derived from existing experimental evidence, and that for radon-neon and radon-argon can be obtained when the thermal diffusion measurements for these two mixtures are completed.

Information concerning the very massive, monatomic radon molecule obtained from diffusion experiments is of special interest, particularly as the molecular field cannot be obtained from viscosity measurements or the equation of state owing to the extremely small quantity of gas available.

EARLY WORK AND PRELIMINARY EXPERIMENTS

Shortly after the discovery of the emanations actinon, thoron and radon, several attempts were made to determine their coefficient of diffusion into air or hydrogen in order to obtain the molecular weights of the emanations. The references to these early investigations are given at the end of this paper.

The most notable feature about the results of these early experiments is that the values obtained for the diffusion coefficient of the isotopes actinon, thoron and radon into air or hydrogen show variations amounting to 25 % or more. It appeared to us that these variations must be due to some common disturbing factor, and the most likely possibility, at first, seemed to be that local differences of temperature existed in the diffusion apparatus which would tend to set up convection in the gas mixture and so disturb the diffusion process.

It was decided to perform a series of diffusion experiments using radon-air and radon-hydrogen mixtures in which special care was taken to avoid temperature variations along the diffusion apparatus. The Loschmidt method was used as successful measurements had recently been made in this laboratory by L. E. Boardman and N. E. Wild (1937) using this method for the determination of the diffusion coefficient of non-radioactive gas mixtures.

The apparatus consisted of a closed brass tube provided with a central tap. The tube was mounted vertically and a glass capillary containing radon was put into the lower half of the apparatus. Hydrogen or air was then introduced and the central tap closed. The radon capillary was broken and the apparatus left for 18 hr. to obtain a uniform gas mixture in the lower half. As the amount of radon was extremely small compared with that of the other gas, the pressure was considered to be equal in the two halves of the apparatus. The central tap was then opened for a known time so that diffusion could take place and the diffusion coefficient was obtained from measurements of the ratio of the amount of radon diffusing over into the upper half to that remaining in the lower half of the apparatus. The gas mixtures used in these experiments were at about atmospheric temperature and pressure.

The results obtained from these preliminary experiments showed variations which were of the same order as those in the results obtained by the early investigators. It was concluded that the cause of these large variations must be due to an experimental difficulty peculiar to radioactive gas mixtures. The relative proportion by volume of radon in the gas mixtures used in our experiments was extremely small (about 1×10^{-6}). The diffusion of radon alone was observed, the other gas having no direct effect upon the measuring instrument. If part of this small proportion of radon was not free to diffuse appreciable variations would be produced in the results obtained.

Further experiments showed that radon adhered to the walls of the brass diffusion tube if it had been in the apparatus for a sufficient time. In an extreme case less than 2 % of the amount of radon put into the lower half of the apparatus diffused over into the upper half, as compared with about 30 % had the diffusion been normal. We have also found that radon adheres to a glass apparatus, which supports the observations made previously by Lind and Livingston (1936). They noticed that if radon was left in pyrex glass bulbs for more than a few hours, appreciable adhesion of the radon to the glass took place.

There is no doubt that this is the principal source of error in our preliminary experiments and those of previous investigators. In the latter, the emanations were often left in the apparatus for several days, and the amount of adhesion, which increases with time, would then become a serious source of error. Since adhesion tends to fix the position of the emanation, it always leads to a low value of the diffusion coefficient. Rutherford and Miss Brooks (1901) obtained successively lower values of the diffusion coefficient for radon-air mixtures, which decreased from 0·15 to 0·08 $\text{cm.}^2 \text{sec.}^{-1}$, the latter value being obtained after using the diffusion apparatus over a period of a year or so.

A series of experiments were made with apparatus of the form already described to determine how adhesion of radon to the walls might be made of negligible account in diffusion experiments. The apparatus had a volume of about 35 c.c. and the surface area was about 50 cm.^2 About 50 mc. of radon was used in each of these experiments, the relative proportion by volume of radon to the other gas thus remaining about 1×10^{-6}. The gas mixtures were varied, radon being mixed with either hydrogen, argon or air. The following conclusions were drawn from these observations:

(1) When radon is enclosed in a brass or glass apparatus which has not previously been in contact with the gas and has not been chemically cleaned, it adheres to the surface of the container. The amount of radon which

adheres to an apparatus of the size already given is negligible during the first few hours after the radon is introduced but reaches 3 or 4 % after about 20 hr. and thereafter increases rapidly. These results are not materially affected by the presence of normal amounts of grease on taps, etc., attached to the apparatus or by the nature of the gas mixed with the radon.

(2) If successive experiments are made using the same apparatus, the rate of adhesion gradually increases in each experiment. The adhesion thus depends on the total time of contact of radon with the apparatus. When about 2 mc. of radon per cm.^{2} have been destroyed in the apparatus, the adhesion in subsequent experiments becomes considerable in a few hours.

(3) If an apparatus which has been contaminated with radon is washed through with strong acid before it is used again, the rate of adhesion is increased. This is most probably due to a roughening of the surface of the apparatus by the acid, and this inference is supported by the observation that radon adheres to a rough brass surface more quickly than to a smooth glass one under similar conditions.

(4) The adhesion forces holding the radon on the surface are weak. This might be expected since radon is monatomic. If the apparatus is evacuated or a stream of air is passed through it most of the radon adhering to the walls is removed.*

(5) If pyrex glass is used and the latter is baked at 600° C after it has contained radon, so that the glass is softened, the original uncontaminated condition is restored and the rate of adhesion follows that given in note (1).

The adhesion of radon must be distinguished from ordinary adsorption. The latter would probably result in a rapid deposition of radon on the walls immediately after the radon was put into the apparatus, and the rate of deposition would then decrease with time. In our experiments there is no measurable adhesion to the walls when the gas is first introduced. This suggests that if ordinary adsorption occurs it is of negligible amount. On the other hand, the rate of adhesion increases rapidly after radon has been in contact with the apparatus for several hours. This would indicate that the adhesion is due to a conditioning of the walls of the apparatus by the radon. This conditioning process is probably brought about by the radioactive properties of the gas. The bombardment by α-particles may result in a change in the condition of the surface so that radon atoms can adhere or the presence of the active deposits on the walls may bring about this change in condition.

---

* It should be noted that in the experiments made by Lind and Livingston (1936) in which the relative proportion by volume of radon was much higher, the adhesion to pyrex glass was more permanent.

The conclusion which may be drawn from these observations so far as the present experiments are concerned is that when radon is used in a diffusion apparatus there is no reliable method of estimating the amount adhering to the walls. As already mentioned, evacuation or even the passage of a current of air through the apparatus removes most of the radon adhering to it. There is no doubt that adhesion is negligible in a glass apparatus which has contained radon for less than 6 hr. If pyrex glass is used in the construction of the apparatus, it should be heated to 600° C periodically to reform the surface. The difficulty of adhesion may thus be overcome.

In our first attempts to avoid prolonged contact of radon with the diffusion apparatus, an arrangement was used for the preliminary mixing in which a radon capillary was broken in a gas burette. After the radon gas mixture had become uniform, it was subsequently introduced into one-half of the diffusion apparatus. This method proved to be an unsuitable one and it did not yield consistent results.

THE FINAL EXPERIMENTAL ARRANGEMENT

(a) *General principle of the method*

A plan of the final form of the diffusion apparatus is shown in fig. 1. It consisted of a uniform pyrex glass tube of length $L = 42{\cdot}7$ cm. and of 6 mm. bore, divided into three equal sections P, Q and R by taps T_1 and T_2 of the same bore.

The apparatus was filled to atmospheric pressure with the gas into which the radon was to diffuse, and the tap T_1 closed. A glass capillary tube containing radon was broken in P and the apparatus was left for 18 hr. to obtain a uniform gas mixture in this section. Tap T_1 was then opened for a known time, so that radon could diffuse over into Q and R, after which the taps T_1 and T_2 were closed simultaneously. The diffusion coefficient was obtained by measuring the ratio of the amounts of radon in sections Q and R after diffusion.

It is important to notice that the measurement of the ratio w/u of the amounts of radon in Q and R need only be obtained, and that the amount of radon in P need not be known, as is shown later in the theory of the experiment. This means that any radon which adhered to the walls in P during the long period of the initial mixing does not affect the results of the experiments. Radon did not come into contact with Q and R until diffusion took place. The diffusion time is relatively short (about $\frac{1}{2}$ hr.), so that adhesion in these sections was negligible. The gas pressure was uniform throughout the apparatus before diffusion commenced, and there were therefore no pressure changes in the sections when the taps were opened.

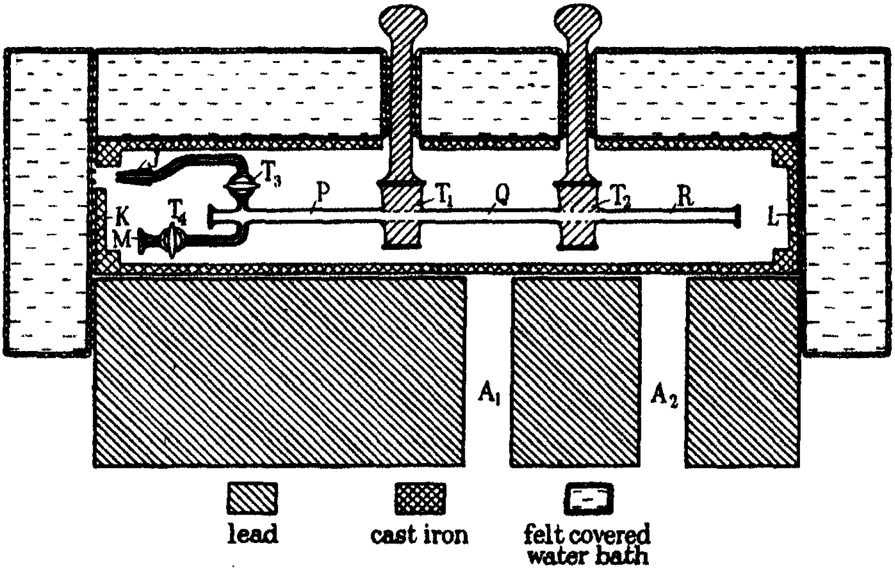


Fig. 1. Plan of apparatus.

The agreement between the results obtained in which the effects of the adhesion of radon are eliminated is as satisfactory as that between the values obtained by Boardman and Wild (1937) for non-radioactive gas mixtures. It should also be mentioned that the results of the thermal diffusion experiments given in the previous investigation (Harrison 1937) are unaffected by adhesion, any experimental values which showed evidence of this being rejected.

(*b*) *Experimental procedure*

As the amount of radon was extremely small, the density of the gas in P, Q and R could be regarded as constant, so that the apparatus could be mounted horizontally thus reducing convection.

A glass capillary tube containing about 70 mc. of radon was pushed into the bore of the capillary tap T_4 and a glass plate M waxed on the end. The apparatus was then connected by the ground joint J to a vacuum pump, manometer, and a cylinder of the gas into which radon was to be diffused. The taps T_1, T_2 and the capillary tap T_3 were then opened and the apparatus exhausted. It was filled with the gas to about atmospheric pressure. Taps

T_1 and T_3 were then closed, tap T_2 remaining open, and the radon capillary was broken in section P by turning T_4.

The diffusion experiment was performed by opening the tap T_1 for a known time which varied from 15 to 40 min., depending on the gas mixing with the radon. Taps T_1 and T_2 were then closed simultaneously. Part of the radon originally in P had now diffused over into Q and R.

As the bores of taps T_1 and T_2 formed part of the diffusion tube, a small side-hole was drilled in the barrel of each of these taps, so that when they were closed the gas in the bore of T_1 was in communication with section Q, and that in the bore of T_2 was in communication with section R. This avoided any error due to gas being isolated in the bore of the taps when the latter were closed.

The ratio of the amount of radon in Q to that in R, after diffusion, was measured by determining the relative intensity of the γ-rays from the solid decay products radium B and C deposited on the walls in Q and R in the process of radioactive change.

A block of lead 22 × 6 × 5 in. was mounted behind the pyrex tube. Two equal rectangular apertures A_1 and A_2 ($1\frac{1}{2} \times \frac{3}{4}$ in.) were cut out of this block, so that the γ-ray intensity from two uniform portions of Q and R could be observed. Four hours after closing the taps T_1 and T_2, when the radon in the sections Q and R was approximately in equilibrium with the γ-ray decay products, a γ-ray electroscope with lead walls 5 mm. thick was placed in turn opposite each of the apertures A_1 and A_2. The ratio of the amounts of radon opposite A_1 and A_2 was obtained by observing the rate of drift of the electroscope when placed opposite each aperture. The rate of drift of the electroscope was also determined when it was moved along the lead block out of line of the apertures, and the above ratio of the rates of drift was thus corrected for the natural leak of the electroscope and for stray γ-radiation.

The sections Q and R were of equal length and cross-sectional area. The radon in each section was uniformly distributed at the time of the radioactive measurements so that, when the apertures A_1 and A_2 are equal, the ratio of the amounts of radon opposite the apertures is equal to the ratio, w/u, of the amounts of radon in the sections Q and R.

In order to allow for any inequality in the effective sizes of the apertures A_1 and A_2, tap T_2 was opened to allow the radon gas mixture in sections Q and R to become uniform. Four hours afterwards, a similar series of measurements of the rate of drift of the electroscope enabled the ratio of the effective sizes of the apertures to be obtained.

Immediately after an experiment was completed the apparatus was exhausted. Radon was therefore in the diffusion apparatus for as short a time as possible. As a precaution the gas apparatus was periodically heated to 600° C.

(c) *Temperature precautions*

To ensure a uniform temperature over the diffusion apparatus the pyrex tube was mounted in a cast iron box provided with a tight-fitting lid (fig. 1). The box was adjacent to the large lead block on one side, and water baths covered with felt were placed round the other three sides. A similar water bath was then placed over the iron box, and a wooden cover was fitted over the apparatus and sealed down to the table with plasticene as a protection against draughts.

The taps T_1 and T_2 were constructed with long handles, so that the ends projected through holes in the iron box and the water baths. They could thus be operated when the temperature lagging was in position, electrical contacts being employed to mark the "on" and "off" positions.

Thermojunctions were used to record any small temperature difference between the two ends of the iron box containing the gas apparatus. One junction from each pair of couples was attached to the inside of the box at K and L, the other junctions being kept in a sealed thermos flask filled with water. These junctions were connected to a recording galvanometer, so that the temperature difference between K and L could be observed for some hours before diffusion commenced as well as while it was in progress. During the diffusion experiments, these differences of temperature did not exceed 0·05° C.

The experiments were performed in a small room provided with thermostatically controlled electric heating.

(d) *The theory of the experiment*

If ρ represents the number of radon molecules per c.c. at any point x along the axis of the tube at a time t after diffusion commences, the differential equation determining the coefficient of diffusion, D_{12}, is

$$\frac{\partial \rho}{\partial t} = D_{12}\frac{\partial^2 \rho}{\partial x^2}. \tag{1}$$

Let L be the total length of the gas apparatus.

The initial conditions (at time $t = 0$) are $\rho = c$, say, from $x = 0$ to $x = L/3$ in section P, while $\rho = 0$ from $x = L/3$ to $x = L$ in sections Q and R (fig. 1).

The boundary conditions are $\frac{\partial\rho}{\partial x} = 0$ at $x = 0$ and $x = L$. Thus

$$\rho = \frac{c}{3} + \frac{2c}{\pi}\sum_{n=1}^{n=\infty}\frac{1}{n}\sin\frac{n\pi}{3}e^{-D_{12}t(n\pi/3)^2}\cos\left(\frac{n\pi}{L}\right)x,$$

where n is a positive integer.

After diffusion let the total quantity of radon in Q be w:

$$w = \int_{L/3}^{2L/3}\rho dx.$$

Similarly, let the total quantity of radon in R be u:

$$u = \int_{2L/3}^{L}\rho dx.$$

Hence $$w = \frac{cL}{9} + \frac{2cL}{\pi^2}\sum_{n=1}^{n=\infty}\frac{1}{n^2}\sin\frac{n\pi}{3}e^{-D_{12}t(n\pi/L)^2}\sin\left[\frac{2n\pi}{3} - \frac{n\pi}{3}\right]$$

and $$u = \frac{cL}{9} + \frac{2cL}{\pi^2}\sum_{n=1}^{n=\infty}\frac{1}{n^2}\sin\frac{n\pi}{3}e^{-D_{12}t(n\pi/L)^2}\left[-\sin\frac{2n\pi}{3}\right].$$

The solution may be expressed in the form

$$\frac{w}{2u+w} = \frac{\frac{1}{9} - \frac{3}{\pi^2}[\frac{1}{4}e^{-4D_{12}t(\pi/L)^2} + \frac{1}{16}e^{-16D_{12}t(\pi/L)^2} + \dots]}{\frac{1}{3} - \frac{3}{\pi^2}[e^{-D_{12}t(\pi/L)^2} - \frac{1}{25}e^{-25D_{12}t(\pi/L)^2} + \dots]},^* \tag{2}$$

in which the series on the right-hand side converges rapidly. This equation may be solved by successive approximation. For the values of $D_{12}t$ occurring in the present experiments only the first approximation was used, since this gave a solution correct to 0·3 %. This first approximation is

$$\frac{w}{2u+w} = \frac{\frac{1}{9}}{\frac{1}{3} - \frac{3}{\pi^2}\{e^{-D_{12}t(\pi/L)^2}\}}.$$

A more convenient form of this first approximation is

$$\frac{u}{w} = 1 - \frac{27}{2\pi^2}e^{-D_{12}t(\pi/L)^2}. \tag{3}$$

It will be seen from equation (3) that the value of D_{12} is independent of c. It is thus unaffected by the amount of radon in P and therefore by the adhesion of radon in this section.

---

[*] Equation (2) is also correct for a radioactive substance, the amount of which decreases exponentially with time.

The time of diffusion, t, must be chosen for each gas mixture so that u and w each represent measurable quantities of radon and yet differ from one another by an amount which can also be determined with accuracy.

(*e*) *Results*

The final experimental results are given in Table I.

TABLE I. FINAL EXPERIMENTAL RESULTS

Gases diffused	Conditions of experiment		Time of diffusion sec.	u/w	D_{12} cm.^{2} sec.$^{-1}$	D_{12} cm.^{2} sec.$^{-1}$ at 15° C and 76 cm. Hg	Mean value of D_{12} cm.^{2} sec.$^{-1}$ at 15° C and 76 cm. Hg
	T° C	Pressure cm. Hg					
Radon-air	11·2	72·5	2400	0·734	0·125	0·121	0·120
Radon-air	10·5	75·7	2400	0·718	0·121	0·123	
Radon-air	13·6	75·9	2400	0·699	0·116	0·117	
Radon-air	10·6	74·8	2400	0·697	0·116	0·117	
Radon-hydrogen	12·4	75·2	610	0·717	0·475	0·471	0·476
Radon-hydrogen	12·4	75·2	600	0·721	0·485	0·481	
Radon-helium	13·0	75·6	600	0·561	0·336	0·340	0·351
Radon-helium	9·5	76·2	600	0·543	0·350	0·362	
Radon-neon	19·4	74·8	1200	0·691	0·228	0·219	0·217
Radon-neon	16·8	76·0	1200	0·668	0·217	0·215	
Radon-argon	9·1	74·7	2700	0·635	0·090	0·091	0·092
Radon-argon	12·4	75·4	2700	0·645	0·092	0·093	

It has already been mentioned that the adhesion of radioactive gas to the walls of a diffusion apparatus leads to an error in the experimental value of D_{12} which makes the latter lower than the true one. The mean value of D_{12} finally obtained for radon air mixtures is about 14 % greater than that obtained in our preliminary experiments or in those of previous investigators. The agreement between the final experimental results given in Table I for a given radon gas mixture is much closer than that obtained hitherto in diffusion experiments using radioactive gas mixtures.

THE CALCULATION OF THE INTERMOLECULAR FIELDS

Chapman (1929) has shown that if c_1 and c_1' represent the proportion by volume of the rare and heavy constituent of a gas mixture at the absolute temperatures T and T' respectively, after thermal diffusion, then

$$\log_e \frac{c_1}{c_1'} = \frac{\alpha}{\beta}\left[\left(\frac{T'}{T}\right)^{\beta} - 1\right], \tag{4}$$

where α is proportional to T^{β}, and for a given pair of gases, β is given by

$$\beta = \frac{1}{2}\left[\frac{s_{12}-5}{s_{12}-1} - \frac{s_2-5}{s_2-1}\right], \tag{5}$$

s_2 being the repulsive force index between like molecules of the lighter gas.

When c_1 and the ratio of the mass of the molecule of the lighter gas to that of the heavy one are both small, a first approximation to α, $\alpha(1)$, is

$$\alpha(1) = \frac{3}{4}\frac{s_{12}-5}{s_{12}-1}\frac{RT}{pM_2}\frac{\eta_2}{D_{12}}, \tag{6}$$

where R is the gas constant per gram molecule, η_2 is the viscosity of the lighter gas at an absolute temperature T, and D_{12} is the coefficient of diffusion of the mixture at any pressure p and absolute temperature T. M_2 is the mass of a gram molecule of the lighter gas.

If η_2 has been determined over a range of temperatures, s_2 may be calculated. The experimental values of D_{12} and the ratio c_1/c_1' obtained from thermal diffusion measurements, then suffice to determine the repulsive force index between dissimilar molecules, s_{12}, using equations (4), (5) and (6).

Measurements of c_1/c_1' have already been made for radon-hydrogen and radon-helium gas mixtures (Harrison 1937). In order to obtain s_{12} for these dissimilar molecules, the value of D_{12} had at that time to be calculated, for each gas mixture, from a formula given by Chapman (1917) in the derivation of which it is assumed that the molecules behave as rigid elastic spheres. The results of the present experiments complete the experimental information necessary to calculate s_{12} for radon-hydrogen and radon-helium molecules.

For hydrogen molecules s_2 was taken as 12·0 and that for helium molecules as 14·6. Substituting the values, $\eta = 89 \times 10^{-6}$ c.g.s. units at 15° C for hydrogen and $\eta = 197 \times 10^{-6}$ c.g.s. units at 15° C for helium, s_{12} for radon-hydrogen molecules was found to be 6·0 and that for radon-helium molecules 6·45.

Chapman (1929) has also shown that a first approximation to D_{12}, $D_{12}(1)$, is given by

$$D_{12}(1) = \frac{3}{16\pi^{\frac{1}{2}}}\frac{M_1+M_2}{M_1M_2}\frac{R^2T^2}{pN\Omega_{12}(11)}, \tag{7}$$

where M_1 and M_2 are the molecular weights of the two constituents of the gas mixture and N is the number of molecules per gram molecule. $\Omega_{12}(11)$ is defined by the equation

$$\Omega_{12}(11) = \frac{1}{4\pi}I_1(s_{12})\,K_{12}^{\frac{2}{s_{12}-1}}\left(2RT\frac{M_1+M_2}{M_1M_2}\right)^{\frac{s_{12}-5}{2(s_{12}-1)}}\Gamma\left(3-\frac{2}{s_{12}-1}\right), \tag{8}$$

where $I_1(s_{12})$ is a constant for a given pair of molecules which has been evaluated for various values of s_{12} and K_{12} is defined by the equation

$$\mu_{12} = K_{12}\frac{m_1 m_2}{m_1 + m_2}, \tag{9}$$

μ_{12} being the repulsive force constant between dissimilar molecules of masses m_1 and m_2 respectively.

When s_{12} is known, the experimental value of D_{12} is sufficient to determine μ_{12} from equations (7), (8) and (9). For radon-hydrogen molecules ($s_{12} = 6{\cdot}0$), $I_1(s_{12})$ is 2·47 and for radon-helium molecules ($s_{12} = 6{\cdot}45$), $I_1(s_{12})$ is 2·45. Thus the value of μ_{12} for radon-hydrogen molecules is $2{\cdot}4 \times 10^{-50}$, and that for radon-helium molecules $9{\cdot}5 \times 10^{-54}$. The repulsive force between these dissimilar molecules may therefore be expressed by the equations

$$F_{12} = 2{\cdot}4 \times 10^{-50} d^{-6{\cdot}0} \text{ radon-hydrogen},$$
$$F_{12} = 9{\cdot}5 \times 10^{-54} d^{-6{\cdot}45} \text{ radon-helium}.$$

Discussion

The repulsive force index for radon-hydrogen and radon-helium molecules which are now derived from experimental evidence, are rather lower than the corresponding approximate values given in Table II of the previous investigation (Harrison 1937). But the revised values of s_{12} do not differ fundamentally from the previous ones and serve to emphasize the extremely "soft" nature of the radon molecule. This is readily seen if the values of s_{12} obtained for radon-hydrogen and radon-helium are compared with those between other unlike pairs of molecules containing hydrogen or helium, given in the table referred to above.

TABLE II. VALUES OF μ_{12} FOR VARIOUS DISSIMILAR MOLECULES

Gas mixture	D_{12} cm.^{2} sec.$^{-1}$ 15° C and 76 cm. Hg	s_{12}	μ_{12}
H_2-N_2	0·748*	8·3†	$6{\cdot}1 \times 10^{-66}$
H_2-CO_2	0·598*	6·9†	$2{\cdot}9 \times 10^{-57}$
H_2-Rn	0·476	6·0	$2{\cdot}4 \times 10^{-50}$
He-A	0·704‡	9·0†	$1{\cdot}2 \times 10^{-74}$
He-Rn	0·351	6·45	$9{\cdot}5 \times 10^{-54}$

* Boardman and Wild (1937). † Ibbs and Grew (1931). ‡ Lonius (1909).

It is also of interest to compare the values of F_{12} obtained for radon-hydrogen and radon-helium with that for other pairs of molecules. The necessary diffusion and thermal diffusion data are available for hydrogen-nitrogen, hydrogen-carbon dioxide and helium-argon mixtures. Using the data for D_{12} and s_{12} which are given in Table II, the value of μ_{12} was calculated for each of these pairs of molecules from equations (7), (8) and (9). These are shown in column 4 of the table, the corresponding values for radon-hydrogen and radon-helium being included.

In each case, when radon is mixed with a particular gas it gives the highest value of μ_{12} for mixtures containing that gas.

Thermal diffusion experiments using radon-neon and radon-argon mixtures are in progress and when these experiments are completed F_{12} may be calculated for these pairs of molecules from the experimental data.

In conclusion, we wish to thank Professor M. L. E. Oliphant, F.R.S., for the full facilities which he has provided for carrying out this work. Our thanks are also due to Dr T. L. Ibbs for his valuable advice and criticism. One of us (W. H.) is also indebted to the Department of Scientific and Industrial Research for a grant to enable this work to be completed.

SUMMARY

Diffusion measurements have been made for gas mixtures containing radon mixed with hydrogen, helium, neon, argon and air. The results of these experiments in the cases of radon-hydrogen and radon-helium mixtures, together with the thermal diffusion experiments already made (Harrison 1937), suffice to determine the repulsive force between these two pairs of molecules.

It is shown that all previous measurements of the coefficient of diffusion of radioactive gas mixtures are subject to an error due to the adhesion of the emanation to the walls of the diffusion apparatus which always tends to reduce the experimental values obtained for the diffusion coefficient. An experimental study has been made of the means by which adhesion can be made of negligible account in diffusion experiments and an apparatus has been designed which eliminates this disturbing effect. The agreement between the final values obtained for the diffusion coefficient of radon gas mixtures is as good as that between the values recently obtained for the coefficient of diffusion of non-radioactive gas mixtures.

The repulsive force between radon-hydrogen and radon-helium molecules is compared with that between hydrogen-nitrogen, hydrogen-carbon dioxide and helium-argon molecules. The radon molecule is shown to be very "soft".

References

Boardman, L. E. and Wild, N. E. 1937 *Proc. Roy. Soc.* A, **162**, 511–20.
Chapman, S. 1917 *Philos. Trans.* A, **217**, 115–97.
— 1929 *Phil. Mag.* 7, 1–16.
Harrison, G. E. 1937 *Proc. Roy. Soc.* A, **161**, 80–94.
Ibbs, T. L. and Grew, K. E. 1931 *Proc. Phys. Soc.* **43**, 142–55.
Lind, S. C. and Livingston, R. 1936 *J. phys. Chem.* **40**, 419–20.
Lonius, A. 1909 *Ann. Phys., Lpz.*, **29**, 664–78.

Early experimental work

Bruhat, G. 1909 *Radium, Paris*, **6**, 67–73.
Chaumont, L. 1909 *Radium, Paris*, **6**, 106–8.
Curie, P. and Danne, J. *C.R. Acad. Sci., Paris*, **136**, 1314–16.
Debierne, A. 1907 *Radium, Paris*, **4**, 213–18.
Leslie, M. 1912 *Phil. Mag.* **24**, 637–47.
McKeehan, L. W. 1917 *Phys. Rev.* **10**, 473–82.
Russ, S. 1909 *Phil. Mag.* **17**, 412–22.
Rutherford, E. and Brooks, H. T. 1901 *Trans. Roy. Soc. Can.* 7, Section 3, 21–25.

INDEX TO VOLUME CLXIX (A)

Address of the President (Bragg), 1.

Adsorption of double molecules, theory (Chang), 512.

Agar (J. N.) and Bowden (F. P.) The kinetics of electrode reactions. I and II, 206.

Aickin (R. G.), Baylis (N. S.) and Rees (A. L. G.) The effect of solvents on the continuous absorption spectrum of bromine, 234.

Airscrew sound waves, photography of (Hilton), 174.

Allibone (T. E.) and Meek (J. M.) The development of the spark discharge. II, 246.

Allmand (A. J.) *See* Lizius and Allmand.

Ammonia in water, basic strength of (Everett and Wynne-Jones), 190.

Ammonium ion, dissociation of (Everett and Wynne-Jones), 190.

Arnot (F. L.) and M'Ewen (M. B.) Electron transfer in mercury vapour, 437.

Bakerian Lecture (Ingold), 149.

Band spectrum of silicon monosulphide (Barrow and Jevons), 45.

Barrow (R. F.) and Jevons (W.) The band spectrum of silicon monosulphide and its relation to the band spectra of similar molecules, 45.

Baylis (N. S.) *See* Aickin and others.

Benzene, structure of (Ingold), 149.

Bowden (F. P.) and Leben (L.) The nature of sliding and the analysis of friction, 371.

Bowden (F. P.) and Tabor (D.) The area of contact between stationary and between moving surfaces, 391.

Bowden (F. P.) *See* Agar and Bowden.

Bradfield (K. N. E.), Christopherson (D. G.) and Southwell (R. V.) Relaxation methods applied to engineering problems. IV. Problems relating to elastic stability and vibrations, 289.

Bragg (Sir William) Address of the President, 1.

Bromine absorption spectrum, effect of solvents on (Aickin and others), 234.

Buckingham (R. A.) *See* Massey and Buckingham.

Cameron (W. H. B.) and Elliott (A.) The emission band spectrum of chlorine. III, 463.

Carbon bisulphide and water vapour, sorption of (Lizius and Allmand), 25.

Chang (T. S.) Statistical theory of the adsorption of double molecules, 512.

Childs (W. H. J.) and Jahn (H. A.) A new Coriolis perturbation in the methane spectrum. III, 451.

Childs (W. H. J.) and Jahn (H. A.) The spectrum of mono-deuteromethane in photographic infra-red. I. Preparation of the gas, and measurements of a simple band at 9021 cm.$^{-1}$, 428.

Christopherson (D. G.) *See* Bradfield and others.

Co-operative assemblies, statistical mechanics of (Guggenheim), 134.

Coriolis perturbation in the methane spectrum (Childs and Jahn), 451.

Cosmic-ray particles, scattering (Williams), 531.

Coulson (C. A.) The electronic structure of some polyenes and aromatic molecules. VII. Bonds of fractional order by the molecular orbital method, 413.
Critical and co-operative phenomena (Lennard-Jones and Devonshire), 317.
Curran (S. C.), Dee (P. I.) and Petržílka (V.) The excitation of X-radiation in process of proton capture by light elements, 269.

Dee (P. I.) *See* Curran and others.
Devonshire (A. F.) *See* Lennard-Jones and Devonshire.
Diffusion of radon gas mixtures (Hirst and Harrison), 573.
Dynamical continuum, Hamilton-Jacobi theory (Weiss), 102.

Elastic stability and vibrations, problems (Bradfield and others), 289.
Electrode reactions, kinetics of (Agar and Bowden), 206.
Electro-dynamics, Hamilton-Jacobi theory, 119.
Electron ferromagnetism (Stoner), 339.
Electron transfer in mercury vapour (Arnot and M'Ewen), 437.
Electrons, fast, scattering (Williams), 531.
Elliott (A.) A band spectrum attributed to NBr, 469.
Elliott (A.) *See* Cameron and Elliott.
Energy and specific heat (Stoner), 339.
Everett (D. H.) and Wynne-Jones (W. F. K.) The dissociation of the ammonium ion and the basic strength of ammonia in water, 190.

Friction, analysis of (Bowden and Leben), 371.

Guggenheim (E. A.) The statistical mechanics of co-operative assemblies, 134.

Hamilton-Jacobi theory (Weiss), 102, 119.
Harrison (G. E.) *See* Hirst and Harrison.
Helium, low temperature properties of (Massey and Buckingham), 205.
Hilton (W. F.) The photography of airscrew sound waves, 174.
Hirst (W.) and Harrison (G. E.) The diffusion of radon gas mixtures, 573.

Ingold (C. K.) The Bakerian Lecture. The structure of benzene, 149.
Ionic recombination in air (Sayers), 83.

Jahn (H. A.) *See* Childs and Jahn.
Jevons (W.) *See* Barrow and Jevons.

Lennard-Jones (J. E.) and Devonshire (A. F.) Critical and co-operative phenomena. III. A theory of melting and the structure of liquids, 317.
Leben (L.) *See* Bowden and Leben.
Lizius (J. L.) and Allmand (A. J.) The simultaneous sorption of carbon bisulphide and water vapour by activated charcoals, 25.

M'Ewen (M. B.) *See* Arnot and M'Ewen.
Massey (H. S. W.) and Buckingham (R. A.) The low temperature properties of gaseous helium, 205.

Meek (J. M.) *See* Allibone and Meek.
Melting, and structure of liquids, a theory of (Lennard-Jones and Devonshire), 317.
Molecular orbital method (Coulson), 413.
Mono-deuteromethane, spectrum in photographic infra-red (Childs and Jahn), 428.

Paraffin wax, polar molecules in (Sillars), 66.
Petržílka (V.) *See* Curran and others.
Polar molecules in solid paraffin wax, behaviour of (Sillars), 66.
Polyenes and aromatic molecules, electronic structure of (Coulson), 413.
Presidential Address (Bragg), 1.
Price (W. C.) and Simpson (D. M.) The absorption spectra of carbon dioxide and carbon oxysulphide in the vacuum ultra-violet, 501.
Proton capture by light elements (Curran and others), 269.

Rees (A. L. G.) *See* Aickin and others.
Relaxation methods applied to engineering problems (Bradfield and others), 289.
Relaxation methods applied to linear systems (Temple), 476.

Sayers (J.) Ionic recombination in air, 83.
Silicon monosulphide, band spectrum of (Barrow and Jevons), 45.
Sillars (R. W.) The behaviour of polar molecules in solid paraffin wax, 66.
Simpson (D. M.) *See* Price and Simpson.
Sliding, the nature of (Bowden and Leben), 371.
Solvents, effect on continuous absorption spectrum of bromine (Aickin and others), 234.
Sorption of carbon bisulphide and water vapour by activated charcoals (Lizius and Allmand), 25.
Southwell (R. V.) *See* Bradfield and others.
Spark discharge, development of (Allibone and Meek), 246.
Spectra, absorption, of carbon dioxide and oxysulphide (Price and Simpson), 502.
Spectrum, band, attributed to NBr (Elliott), 469.
Spectrum of chlorine, emission band (Cameron and Elliott), 463.
Stoner (E. C.) Collective electron ferromagnetism. II. Energy and specific heat, 339.
Surfaces, moving and stationary, area of contact between (Bowden and Tabor), 391.

Tabor (D.) *See* Bowden and Tabor.
Temple (G.) The general theory of relaxation methods applied to linear systems, 476.

Weiss (P.) On the Hamilton-Jacobi theory and quantization of a dynamical continuum, 102.

Weiss (P.) On the Hamilton-Jacobi theory and quantization of generalised electro-dynamics, 119.
Williams (E. J.) Concerning the scattering of fast electrons and of cosmic-ray particles, 531.
Wynne-Jones (W. F. K.) *See* Everett and Wynne-Jones.

X-radiation, excitation of (Curran and others), 269.

END OF THE ONE HUNDRED AND SIXTY-NINTH VOLUME (SERIES A)